ENERGETICALLY CONSISTENT MODEL REDUCTION FOR METRIPLECTIC SYSTEMS

ANTHONY GRUBER[1,*], MAX GUNZBURGER[1], LILI JU[2], AND ZHU WANG[2]

Abstract. The metriplectic formalism is useful for describing complete dynamical systems which conserve energy and produce entropy. This creates challenges for model reduction, as the elimination of high-frequency information will generally not preserve the metriplectic structure which governs long-term stability of the system. Based on proper orthogonal decomposition, a provably convergent metriplectic reduced-order model is formulated which is guaranteed to maintain the algebraic structure necessary for energy conservation and entropy formation. Numerical results on benchmark problems show that the proposed method is remarkably stable, leading to improved accuracy over long time scales at a moderate increase in cost over naive methods.

Keywords: model reduction, metriplectic dynamics, GENERIC formalism, Hamiltonian systems

1. Introduction

Metriplectic dynamical systems offer a prototypical example of how the algebraic structure internal to a system can govern the behavior of its observable quantities. Also referred to as GENERIC systems (see [1]), metriplectic dynamics are produced through a combination of reversible and irreversible contributions whose constituent parts are a noncanonical Poisson structure and a degenerate Riemannian metric structure, respectively. Mathematically, these structures are reflected by algebraic brackets which formally separate the dynamics into terms that are "energy-preserving" and terms that are "dissipative" (see Section 1.1). Combined with appropriate compatibility (or degeneracy) conditions, this seemingly simple idea is geometrically rich and encodes a strong form of the first and second laws of thermodynamics, making it powerful enough to represent many physical systems of interest (see Section 1.2). On the other hand, standard computationally efficient reduced-order models (ROMs) for these systems based on purely statistical considerations will not generally preserve the rich structure afforded by the metriplectic formalism, which can lead to unreasonable or unrealistic results in real-time use cases (see e.g. Section 5). To elucidate the benefits of metriplectic structure-preservation in the context of model reduction, a genuinely metriplectic ROM based on proper orthogonal decomposition (POD) is proposed in Theorem 3.4 which is shown in Theorem 4.2 to converge to the true solution as the reduced dimension increases. The remainder of the manuscript is dedicated to a detailed description of this ROM along with an evaluation of its performance on benchmark examples.

1.1. Overview.

It is first useful to recall metriplectic systems in more detail. The generator for metriplectic dynamics is a notion of free energy $F = E + S$ described by functions $E, S : \mathcal{P} \to \mathbb{R}$ (representing energy and entropy, respectively) which are defined on some phase space $\mathcal{P}$ that may be finite or infinite dimensional. In this case, any observable quantity $\mathbf{O} : \mathcal{P} \to M \subset \mathbb{R}^N$ (for some N) evolves as

$$\dot{\mathbf{O}} = \{\mathbf{O}, F\} + [\mathbf{O}, F] = \{\mathbf{O}, E\} + [\mathbf{O}, S]\,,$$

where $\{\cdot,\cdot\}$ is a noncanonical Poisson bracket on $\mathcal{P}$ capturing the reversible dynamics and $[\cdot,\cdot]$ is a degenerate metric bracket on $\mathcal{P}$ capturing the irreversible dynamics. Metriplectic structure is enforced by the implicit degeneracy conditions $\{S,\cdot\} = [E,\cdot] = \mathbf{0}$, which guarantee an analogue of energy conservation and entropy production. To describe this more precisely, recall that the Poisson structure $\{\cdot,\cdot\}$ is a Lie algebra realization on functions and so is bilinear and skew-symmetric (SS), while the degenerate metric structure $[\cdot,\cdot]$ is chosen to be bilinear and symmetric positive semi-definite (SPSD). This allows for concrete expression of

*Corresponding author: (Anthony Gruber) anthony.gruber@fsu.edu.

the reversible and irreversible brackets as

$$\{\mathbf{O}, E\} = \nabla \mathbf{O} \cdot \mathbf{L} \nabla E,$$
$$[\mathbf{O}, S] = \nabla \mathbf{O} \cdot \mathbf{M} \nabla S,$$

where $\cdot$ represents a choice of inner product on $\mathcal{P}$, ∇ denotes the gradient with respect to $\cdot$ defined through $dF(\mathbf{v}) = \nabla F \cdot \mathbf{v}$, and $\mathbf{L}, \mathbf{M} : \mathcal{P} \to \mathcal{P}$ are SS resp. SPSD linear operators which may depend on the state $\mathbf{x}$. Here again no distinction is made between finite and infinite dimensional systems, as this affects only the choice of inner product $\cdot$. In many cases of interest the observable $\mathbf{O}(\mathbf{x}) = \mathbf{x}$ is simply the identity, so that the system above further simplifies to the standard equations for metriplectic dynamics [2, 1],

$$\dot{\mathbf{x}} = \{\mathbf{x}, E\} + [\mathbf{x}, S] = \mathbf{L} \nabla E + \mathbf{M} \nabla S, \tag{1}$$

which in view of the compatibility conditions

$$\mathbf{L} \nabla S = \mathbf{M} \nabla E = \mathbf{0}, \tag{2}$$

preserve a strong form of the first and second thermodynamical laws. In particular, since $\mathbf{L}^\intercal = -\mathbf{L}$,

$$\dot{E} = \dot{\mathbf{x}} \cdot \nabla E = \mathbf{L} \nabla E \cdot \nabla E + \mathbf{M} \nabla S \cdot \nabla E = \nabla S \cdot \mathbf{M} \nabla E = 0,$$

so that the energy E is conserved along the evolution. Similarly, the fact that $\mathbf{M}^\intercal = \mathbf{M}$ is SPSD implies the relationship

$$\dot{S} = \dot{\mathbf{x}} \cdot \nabla S = \mathbf{L} \nabla E \cdot \nabla S + \mathbf{M} \nabla S \cdot \nabla S = -\nabla E \cdot \mathbf{L} \nabla S + \mathbf{M} \nabla S \cdot \nabla S = |\nabla S|_M^2 \geq 0,$$

so that the entropy S is nondecreasing. Here it becomes clear that asymptotic stability is built-in to the metriplectic framework, as choosing $-S$ as a Lyapunov function shows that solutions to (1) will naturally relax to the state $\nabla F = \mathbf{0}$. Moreover, this gives a degree of freedom in describing a physical system with metriplectic structure, as S can be chosen judiciously from the Casimirs of the Poisson bracket i.e. those functions which annihilate it. Geometrically, it is interesting to observe that the motion of $\mathbf{x}$ is everywhere tangent to the level curves of E and transverse to those of S, which is reflective of the fact that metriplectic dynamics are a combination of Hamiltonian and generalized gradient flows. When $\mathbf{M} = \mathbf{0}$, $E = H$ is the Hamiltonian function, and $\mathbf{L} = \mathbf{J}$ is a square root of $-\mathbf{I}$ (note the freedom in sign), (1) reduces to Hamilton's equations of motion $\dot{\mathbf{x}} = \{\mathbf{x}, H\} = \mathbf{J} \nabla H$. Similarly, when $\mathbf{L} = \mathbf{0}$ and $S = -G$ for some $G : \mathcal{P} \to \mathbb{R}$, (1) reduces to a generalized gradient flow $\dot{\mathbf{x}} = -[\mathbf{x}, G] = -\mathbf{M} \nabla G$.

1.2. **Related Work.** The metriplectic/GENERIC forms of many physical systems have already been proposed and studied theoretically for some time. The compressible Navier-Stokes equations were seen to be metriplectic in [2], and general complex fluids were incorporated into the formalism in [3]. This paved the way for the inclusion of other physical phenomena such as those based on Korteweg-type fluids [4] and the Smoluchowski equation for colloidal suspensions [5]. Moreover, a constrained GENERIC rheological model for polymer solutions was developed in [6] and shown to be effective in predicting steady shear viscosity, while a formulation of dissipative magnetohydrodynamics was discovered in [7] and used in studying two-dimensional incompressible plasma flow. Beyond fluids, metriplectic structure has also been useful in describing mechanical systems such as three-dimensional rigid body dynamics [8], Hamiltonian systems with friction [9], a Vlasov-Fokker-Planck equation [10], and others based on large deviation principles in physics.

There have been far fewer works addressing the computational aspects of metriplectic systems, though some noteworthy progress has been made. Structure preserving numerical methods for finite strain thermoelastodynamics in GENERIC form are discussed in [11], where so-called Energy-Momentum-Entropy consistent schemes are shown to increase stability of the discrete system. A compatible discretization for GENERIC problems using finite elements in space and a monolithic integrator in time was developed in [12] and applied to nonlinear problems in thermoelasticity, again demonstating improved stability properties. There is also a promising line of research into metriplectic integrators using neural network technology, which has produced works such as [13, 14].

From the perspective of model reduction, it has long been recognized that computational models perform better when informed by the algebraic structure of the systems that they are modeling. This remains true for low-dimensional approximation, where the model being approximated is itself a surrogate for some physical phenomena. This has produced an entire subfield of structure-preserving model reduction, whose goal is to design effective low-fidelity surrogates which preserves desired properties of the high-fidelity model under consideration. The relative ubiquity and rich mathematical structure of Hamiltonian systems has inspired

several works on Hamiltonian structure-preserving ROM [15, 16, 17, 18, 19], as well as numerous extensions to the port-Hamiltonian formalism [20, 21, 22, 23, 24] which is a useful generalization of its namesake where the Hamiltonian structure on the interior is allowed to interface with general "ports" on the boundary.

Remark 1.1. In fact, the system (1) can be embedded into the port-Hamiltonian formalism,

$$\begin{aligned}\dot{\mathbf{x}} &= (\mathbf{J}-\mathbf{R})\nabla H(\mathbf{x}) + \mathbf{B}\mathbf{u}(t),\\ \mathbf{y} &= \mathbf{B}^\intercal \nabla H(\mathbf{x}),\end{aligned}$$

where $\mathbf{J}$ is SS and $\mathbf{R}$ is SPSD. In particular, decompose $\mathbf{M} = \mathbf{C}\mathbf{D}\mathbf{C}^\intercal$. Then, choosing $\mathbf{R} = \mathbf{0}$, $\mathbf{J} = \mathbf{L}$ $\mathbf{B} = \mathbf{C}, \mathbf{u} = -\mathbf{D}\mathbf{y}$, and $H = E - S$ the exegy function of the system, it follows that

$$\dot{\mathbf{x}} = \mathbf{L}\left(\nabla E - \nabla S\right) - \mathbf{C}\mathbf{D}\mathbf{C}^\intercal\left(\nabla E - \nabla S\right) = \mathbf{L}\nabla E + \mathbf{M}\nabla S,$$

since $\mathbf{L}\nabla S = \mathbf{M}\nabla E = \mathbf{0}$. On the other hand, none of the port-Hamiltonian ROM work known to the authors can guarantee preservation of the degeneracy conditions (2) necessary for metriplectic structure.

Apart from Hamiltonian systems and their extensions, significant work involving structure-preserving ROM has also been done on topics such as moment-preserving Krylov subspace projection [25] and Lagrangian variational problems [26]. An interpolatory model reduction strategy preserving symmetry, higher order structure, and state constraints is discussed in [27], and a ROM for damped wave propagation in transport networks is developed in [28]. It is remarkable that the strategy in [28] is similar to ours in that the preservation of algebraic compatibility conditions at the reduced level assures desired properties such as conservation of mass, dissipation of energy, passivity, and existence of steady states at the full resolution.

2. Preliminaries

To describe the present method for metriplectic model reduction, it is useful to review some basics of POD-based ROMs. First, recall that the goal is to study systems which conserve some notion of energy E, so it is beneficial to express any approximation $\tilde{\mathbf{x}} \approx \mathbf{x} \in \mathbb{R}^N$ to the full-order state as a perturbation from some reference configuration $\mathbf{x}_0 \in \mathbb{R}^N$, i.e. $\tilde{\mathbf{x}} = \mathbf{x}_0 + \mathbf{U}\hat{\mathbf{x}}$ where $\hat{\mathbf{x}} \in \mathbb{R}^n$ and $\mathbf{U} : \mathbb{R}^n \to \mathbb{R}^N$. This ensures the true value of E is exactly preserved at least at the point where $\hat{\mathbf{x}} = \mathbf{0}$, which serves as the initial condition for the reduced-order system.

Consider the standard POD-ROM procedure with this in mind. Let $\mathbf{x} \in \mathbb{R}^N$ be a semi-discrete object representing the solution to a system of $N \in \mathbb{R}$ ODEs, and let $\mathbf{Y} \in \mathbb{R}^{N\times n_t}$ be a matrix with rank $r \leq \min\{N, n_t\}$ containing snapshots of the high-fidelity solution $\mathbf{w} = \mathbf{x} - \mathbf{x}_0$ at n_t discrete points in the interval $[0, T]$ where $T \in \mathbb{R}$ represents the final simulation time. If $\mathbf{Y} = \tilde{\mathbf{U}}\mathbf{\Sigma}\mathbf{V}^\intercal$ is the singular value decomposition, standard computations show that the matrix $\mathbf{U} \in \mathbb{R}^{N\times n}$ comprised of the first $n < r$ columns of $\tilde{\mathbf{U}}$ minimizes the $L^2([0, T])$ reconstruction error of $\mathbf{w}$, and that this error is precisely the sum of the remaining squared singular values [29]. More precisely, it follows that

$$\left\|\mathbf{w} - \mathbf{U}\mathbf{U}^\intercal\mathbf{w}\right\|^2 := \int_0^T \left|\mathbf{w} - \mathbf{U}\mathbf{U}^\intercal\mathbf{w}\right|^2 dt = \sum_{i=n+1}^{r} \sigma_i^2,$$

where σ_i is the i^{th} singular value of $\mathbf{Y}$. This is the basis for the standard POD-ROM procedure, which is applied to the system governing $\mathbf{x}$ by making the approximation $\tilde{\mathbf{x}} = \mathbf{x}_0 + \mathbf{U}\hat{\mathbf{x}}$ and using that $\mathbf{U}^\intercal\mathbf{U} = \mathbf{I}$ in $\mathbb{R}^n$. In the case of the metriplectic system (1), this yields the reduced-order model

$$\dot{\hat{\mathbf{x}}} = \mathbf{U}^\intercal\mathbf{L}(\tilde{\mathbf{x}})\nabla E(\tilde{\mathbf{x}}) + \mathbf{U}^\intercal\mathbf{M}(\tilde{\mathbf{x}})\nabla S(\tilde{\mathbf{x}}), \tag{3}$$

which is the system of n scalar ODEs that best approximates the FOM (1) in the above sense, but clearly does not preserve the compatibility conditions (2) necessary for metriplectic structure. As will be seen in the numerical experiments (see Section 5), this creates instability which can lead to unphysical blow-up of the solution in time.

Remark 2.1. For notational convenience, dependence on the states $\mathbf{x}, \tilde{\mathbf{x}}, \hat{\mathbf{x}}$ is suppressed when the context is clear. Similarly, the Einstein summation convention is adopted so that any tensor index appearing both up and down in an expression is summed over its appropriate range.

As a first attempt at remedying this, it is reasonable to consider searching for mappings $\hat{\mathbf{L}}, \hat{\mathbf{M}}$ depending only on $\hat{\mathbf{x}}$ such that

$$\mathbf{U}^\intercal \mathbf{L} = \hat{\mathbf{L}} \mathbf{U}^\intercal, \qquad \mathbf{U}^\intercal \mathbf{M} = \hat{\mathbf{M}} \mathbf{U}^\intercal. \tag{4}$$

This would convert (3) into the best possible ROM,

$$\dot{\hat{\mathbf{x}}} = \mathbf{U}^\intercal \mathbf{L}(\tilde{\mathbf{x}}) \nabla E(\tilde{\mathbf{x}}) + \mathbf{U}^\intercal \mathbf{M}(\tilde{\mathbf{x}}) \nabla S(\tilde{\mathbf{x}}) = \hat{\mathbf{L}}(\hat{\mathbf{x}}) \nabla \hat{E}(\hat{\mathbf{x}}) + \hat{\mathbf{M}}(\hat{\mathbf{x}}) \nabla \hat{S}(\hat{\mathbf{x}}),$$

where we have introduced the notation

$$\hat{F} = F \circ \tilde{\mathbf{x}}, \qquad \nabla \hat{F} = \tilde{\mathbf{x}}' \cdot \nabla F = \mathbf{U}^\intercal \nabla F.$$

Note that the compatibility conditions (2) are automatically satisfied in this case, as $\hat{\mathbf{L}} \nabla \hat{S} = \hat{\mathbf{L}} \mathbf{U}^\intercal \nabla S = \mathbf{U}^\intercal \mathbf{L} \nabla S = \mathbf{0}$ and similarly for $\hat{\mathbf{M}} \nabla \hat{E}$. On the other hand, (4) is an overdetermined system of equations when $N > n$, and solving the normal equations gives only the system

$$\hat{\mathbf{x}} = \hat{\mathbf{L}} \nabla \hat{E} + \hat{\mathbf{M}} \nabla \hat{S}, \qquad \hat{\mathbf{L}} = \mathbf{U}^\intercal \mathbf{L} \mathbf{U}, \qquad \hat{\mathbf{M}} = \mathbf{U}^\intercal \mathbf{M} \mathbf{U}. \tag{5}$$

which has the advantage of informing the low-dimensional system with the symmetry relationships $\hat{\mathbf{L}}^\intercal = -\hat{\mathbf{L}}$ and $\hat{\mathbf{M}}^\intercal = \hat{\mathbf{M}}$, but still cannot guarantee metriplectic structure preservation. Since $\mathbf{U}\mathbf{U}^\intercal \neq \mathbf{I}$, this gives only

$$\hat{\mathbf{L}} \nabla \hat{S} = \mathbf{U}^\intercal \mathbf{L} \mathbf{U} \mathbf{U}^\intercal \nabla S \neq \mathbf{0},$$
$$\hat{\mathbf{M}} \nabla \hat{E} = \mathbf{U}^\intercal \mathbf{M} \mathbf{U} \mathbf{U}^\intercal \nabla E \neq \mathbf{0},$$

so that the first and second laws of thermodynamics remain violated. Note that in the case $\mathbf{L} = \mathbf{0}$ or $\mathbf{M} = \mathbf{0}$, the conditions (2) are vacuous and (5) provides a useful reduced order model which preserves some structure present in the original system. In fact, this ROM in the case $\mathbf{M} = \mathbf{0}$ is precisely the Hamitonian structure-preserving ROM proposed in [16].

2.1. **Metriplectic Structure Preservation.** One way to preserve metriplectic structure is to incorporate the compatibility conditions (2) explicitly. Let $\mathbf{e}_k$ denote the k^{th} standard basis vector in $\mathbb{R}^N$ and denote $\nabla F = F^k \mathbf{e}_k$ where $F^k = \mathbf{e}_k \cdot \nabla F = \partial F / \partial x^k$. Consider solving the underdetermined equations (for each $1 \leq i, j \leq N$)

$$\begin{aligned} L_{ij} &= \xi_{ijk} S^k, \\ M_{ij} &= \zeta_{ikjl} E^k E^l, \end{aligned} \tag{6}$$

for tensors $\boldsymbol{\xi}, \boldsymbol{\zeta}$ in $\left(\mathbb{R}^N\right)^{\otimes 3}, \left(\mathbb{R}^N\right)^{\otimes 4}$ respectively which may depend on the state $\mathbf{x}$ and which satisfy the symmetry relations

$$\begin{aligned} \xi_{ijk} &= -\xi_{jik} = -\xi_{ikj}, \\ \zeta_{ikjl} &= -\zeta_{kijl} = -\zeta_{iklj} = \zeta_{jlik}. \end{aligned} \tag{7}$$

This is always possible as long as $\nabla E, \nabla S$ are nonzero for all $\mathbf{x}$ (c.f. Proposition 3.1), and otherwise the metriplectic structure is degenerate. Notice that (7) is simply the coordinate-wise expression of total antisymmetry in $\boldsymbol{\xi}$ as well as symmetric $12-34$ pairwise antisymmetry in $\boldsymbol{\zeta}$. The advantage of this approach is that the degeneracy conditions now follow immediately from the symmetries (7). Identifying $\mathbb{R}^N$ with its dual to make use of the canonical "index-raising" isomorphism, for any $1 \leq i \leq N$ it follows that

$$\left(\mathbf{L} \nabla S\right)^i = \xi^i_{jk} S^k S^j = \xi^i_{kj} S^j S^k = -\xi^i_{jk} S^j S^k = 0,$$
$$\left(\mathbf{M} \nabla E\right)^i = \zeta^i_{kjl} E^k E^l E^j = -\zeta^i_{kjl} E^k E^l E^j = 0,$$

since in either case there is a contraction of the same vector over an antisymmetric pair of indices. Therefore, if reduced-order objects $\hat{\boldsymbol{\xi}}, \hat{\boldsymbol{\zeta}}$ which preserve (7) can be found, the degeneracy conditions (2) will hold by construction.

To that end, notice that (1) and (3) can be rewritten respectively as

$$\dot{\mathbf{x}} = \boldsymbol{\xi}(\nabla S) \nabla E + \boldsymbol{\zeta}(\nabla E, \nabla E) \nabla S,$$
$$\dot{\hat{\mathbf{x}}} = \mathbf{U}^\intercal \boldsymbol{\xi}(\nabla S) \nabla E + \mathbf{U}^\intercal \boldsymbol{\zeta}(\nabla E, \nabla E) \nabla S,$$

where the tensors $\boldsymbol{\xi}, \boldsymbol{\zeta}$ are written suggestively to indicate their role as matrix-valued mappings. This encourages the search for reduced-order matrices

$$\hat{\mathbf{L}} = \hat{\boldsymbol{\xi}}\left(\nabla \hat{S}\right)$$
$$\hat{\mathbf{M}} = \hat{\boldsymbol{\zeta}}\left(\nabla \hat{E}, \nabla \hat{E}\right),$$

defined in terms of reduced tensors $\hat{\boldsymbol{\xi}}, \hat{\boldsymbol{\zeta}}$ which satisfy

$$\begin{aligned} \mathbf{U}^\intercal \boldsymbol{\xi} &= \hat{\boldsymbol{\xi}}\left(\mathbf{U}^\intercal\right)\mathbf{U}^\intercal, \\ \mathbf{U}^\intercal \boldsymbol{\zeta} &= \hat{\boldsymbol{\zeta}}\left(\mathbf{U}^\intercal, \mathbf{U}^\intercal\right)\mathbf{U}^\intercal. \end{aligned} \tag{8}$$

This is generally impossible when $N > n$ for the same reason as before, but (8) can again be interpreted as normal equations whose solutions yield $\hat{\boldsymbol{\xi}} = \mathbf{U}^\intercal \boldsymbol{\xi}\left(\mathbf{U}\right)\mathbf{U}$ and $\hat{\boldsymbol{\zeta}} = \mathbf{U}^\intercal \boldsymbol{\zeta}\left(\mathbf{U}, \mathbf{U}\right)\mathbf{U}$. It is straightforward to check that $\hat{\boldsymbol{\xi}}, \hat{\boldsymbol{\zeta}}$ computed this way satisfy the necessary symmetries (7), in which case the ROM

$$\dot{\hat{\mathbf{x}}} = \left\{\hat{\mathbf{x}}, \nabla \hat{E}\left(\hat{\mathbf{x}}\right)\right\} + \left[\hat{\mathbf{x}}, \nabla \hat{S}\left(\hat{\mathbf{x}}\right)\right] := \hat{\mathbf{L}}(\tilde{\mathbf{x}})\nabla \hat{E}\left(\hat{\mathbf{x}}\right) + \hat{\mathbf{M}}(\tilde{\mathbf{x}})\nabla \hat{S}\left(\hat{\mathbf{x}}\right), \tag{9}$$

will preserve the original metriplectic structure by construction. Suppressing dependence on the state, it follows from symmetry considerations as before that

$$\hat{\mathbf{L}}\nabla \hat{S} = \hat{\boldsymbol{\xi}}\left(\nabla \hat{S}\right)\nabla \hat{S} = \mathbf{0},$$
$$\hat{\mathbf{M}}\nabla \hat{E} = \hat{\boldsymbol{\zeta}}\left(\nabla \hat{E}, \nabla \hat{E}\right)\nabla \hat{E} = \mathbf{0},$$

so that the first and second laws of thermodynamics become

$$\dot{\hat{E}} = \dot{\hat{\mathbf{x}}} \cdot \nabla \hat{E} = \hat{\mathbf{L}}\nabla \hat{E} \cdot \nabla \hat{E} + \nabla \hat{S} \cdot \hat{\mathbf{M}}\nabla \hat{E} = 0,$$
$$\dot{\hat{S}} = \dot{\hat{\mathbf{x}}} \cdot \nabla \hat{S} = -\nabla \hat{E} \cdot \hat{\mathbf{L}}\nabla \hat{S} + \hat{\mathbf{M}}\nabla \hat{S} \cdot \nabla \hat{S} = \left|\nabla \hat{S}\right|^2_{\hat{M}} \geq 0,$$

as desired.

Remark 2.2. It should be mentioned that the "metriplectic-preserving" moniker used to describe (9) is a slight abuse of terminology in the case $\mathbf{L} = \mathbf{L}(\mathbf{x})$, as the reduced Poisson bracket $\{\cdot, \cdot\}$ generated by $\hat{\boldsymbol{\xi}} = \hat{\boldsymbol{\xi}}(\mathbf{x})$ is not guaranteed to satisfy the Jacobi identity for arbitrary $F, G, H : \mathbb{R}^n \to \mathbb{R}$,

$$\{F, \{G, H\}\} + \{G, \{H, F\}\} + \{H, \{F, G\}\} = \mathbf{0}.$$

In fact, a direct calculation shows that the above Jacobi identity is equivalent to the statement (sum on l)

$$\hat{L}_{il}\hat{L}_{jk,l} + \hat{L}_{jl}\hat{L}_{ki,l} + \hat{L}_{kl}\hat{L}_{ij,l} = 0, \qquad 1 \leq i, j, k \leq n,$$

which may not vanish if $\mathbf{L}(\mathbf{x})$ is state-dependent. Conversely, the same calculation shows that the ROM (9) is truly metriplectic outside of this case, since the terms involving second derivatives cancel solely due to symmetry properties.

3. Computing the Metriplectic ROM

To make use of the metriplectic ROM (9), it is necessary to have a reasonable way to compute $\boldsymbol{\xi}, \boldsymbol{\zeta}$ and their reduced-order counterparts $\hat{\boldsymbol{\xi}}, \hat{\boldsymbol{\zeta}}$. Note that it is prohibitively expensive to compute general 3^{rd} and 4^{th}-order tensors online even for moderately large N, as the number of tensor entries is exponential in the degree. Therefore, it is necessary to find an efficient way to express (9) which does not require the explicit construction of these tensors. To that end, recall that $\mathbf{M}$ is SPSD and so has the eigenvalue decomposition

$$\mathbf{M} = \sum_{\alpha=1}^{r} \lambda_\alpha \, \mathbf{m}^\alpha \otimes \mathbf{m}^\alpha,$$

where $1 \leq r \leq N$ and all $\lambda_\alpha > 0$ are positive. If it is possible to write $\mathbf{m}^\alpha = \mathbf{A}^\alpha \nabla E$ for some SS matrices $\mathbf{A}^\alpha$, the tensor

$$\boldsymbol{\zeta} = \sum_{\alpha=1}^{r} \lambda_\alpha \, \mathbf{A}^\alpha \otimes \mathbf{A}^\alpha,$$

would satisfy the desired conditions (6) and (7). The next result shows that this can always be done for metriplectic systems.

Proposition 3.1. *Let k_0, k_1 be indices such that $E^{k_0} \neq 0$ and $S^{k_1} \neq 0$. Suppose $\mathbf{M} = \sum_\alpha \lambda_\alpha \, \mathbf{m}^\alpha \otimes \mathbf{m}^\alpha$ and $\mathbf{B}, \mathbf{C}$ are tensors such that $B^\alpha_{ik_0} = v^\alpha_i / E^{k_0}$, $C_{ijk_1} = L_{ij}/S^{k_1}$, and $B^\alpha_{ik} = C_{ijk} = 0$ otherwise. Then, the tensors $\boldsymbol{\xi}$ and $\boldsymbol{\zeta} = \sum_\alpha \lambda_\alpha \, \mathbf{A}^\alpha \otimes \mathbf{A}^\alpha$ with components*

$$\begin{aligned} \xi_{ijk} &= \frac{1}{2}\left(C_{ijk} + C_{jki} + C_{kij} - C_{jik} - C_{kji} - C_{ikj}\right), \\ A^\alpha_{ik} &= B^\alpha_{ik} - B^\alpha_{ki}, \end{aligned}$$

satisfy (6) *and* (7).

Proof. This is a direct consequence of the compatibility conditions $\mathbf{L}\nabla S = \mathbf{M}\nabla E = \mathbf{0}$. More precisely, Let $\mathbf{B}, \mathbf{C}$ be as in the statement of the Proposition. First, recall that all eigenvectors $\mathbf{m}^\alpha$ of $\mathbf{M}$ are linearly independent with positive eigenvalues $\lambda_\alpha > 0$. The compatibility condition $\mathbf{M}\nabla E = \mathbf{0}$ then becomes

$$\mathbf{M}\nabla E = \sum_{\alpha=1}^{r} \lambda_\alpha \left(\mathbf{m}^\alpha \cdot \nabla E\right) \mathbf{m}^\alpha = \mathbf{0},$$

so it follows that $\mathbf{m}^\alpha \cdot \nabla E = 0$ for all $1 \leq \alpha \leq r$. Using $[F]$ to denote the indicator function of the statement F, the definition of $\mathbf{A}$ then implies that for all α, i,

$$A^\alpha_{ik} E^k = \left(B^\alpha_{ik} - B^\alpha_{ki}\right) E^k = m^\alpha_i - [i = k_0] \, B^\alpha_{ki} E^k = m^\alpha_i - [i = k_0] \, \frac{\mathbf{m}^\alpha \cdot \nabla E}{E^{k_0}} = m^\alpha_i,$$

which establishes that $\mathbf{A}^\alpha \nabla E = \mathbf{m}^\alpha$ for all α. It follows that for all $1 \leq i, j \leq N$,

$$\zeta_{ikjl} E^k E^l = \sum_{\alpha=1}^{r} \lambda_\alpha A^\alpha_{ik} A^\alpha_{jl} E^k E^l = \sum_{\alpha=1}^{r} \lambda_\alpha m^\alpha_i m^\alpha_j = M_{ij},$$

which establishes the second part of (6). The corresponding symmetry relationship in (7) follows immediately from the skew-symmetry of $\mathbf{A}^\alpha$ and the definition of $\boldsymbol{\zeta}$ as a sum of symmetric products. Moving to the case of $\boldsymbol{\xi}$, it is straightforward to compute

$$\begin{aligned} 2\xi_{ijk} E^k &= \left(C_{ijk} + C_{jki} + C_{kij} - C_{jik} - C_{kji} - C_{ikj}\right) S^k \\ &= \left(L_{ij} - L_{ji}\right) + [i = k_1] \left(C_{jki} - C_{kji}\right) S^k + [j = k_1] \left(C_{kij} - C_{ikj}\right) S^k \\ &= 2\left(L_{ij} + [i = k_1] \, \frac{\left(\mathbf{L}\nabla S\right)_j}{S^{k_1}} - [j = k_1] \, \frac{\left(\mathbf{L}\nabla S\right)_i}{S^{k_1}}\right) = 2L_{ij}, \end{aligned}$$

which follows since $\mathbf{L}\nabla S = \mathbf{0}$. This establishes (6), and the antisymmetry condition (7) is immediate since $\boldsymbol{\xi}$ is a multiple of the antisymmetrization of $\mathbf{C}$. □

Remark 3.1. It is interesting to note that a weaker form of Proposition 3.1 remains true in the general case of any 4-tensor $\boldsymbol{\zeta}$ satisfying the symmetries (7). In particular, it follows that $\boldsymbol{\zeta}$ decomposes as the sum of at most N^2 outer products with antisymmetric factors. This is a consequence of a simple reshaping argument: Consider the column-wise unfolding of $\boldsymbol{\zeta}$, denoted $\bar{\boldsymbol{\zeta}}$ and defined componentwise as

$$\bar{\zeta}_{(k-1)N+i,(l-1)N+j} = \zeta_{ikjl},$$

which is an $N^2 \times N^2$ matrix that is symmetric by the 12-34 interchange symmetry of $\boldsymbol{\zeta}$. Its spectral decomposition is

$$\bar{\boldsymbol{\zeta}} = \sum_{\alpha=1}^{s} \mu_\alpha \, \mathbf{w}^\alpha \otimes \mathbf{w}^\alpha,$$

where $1 \leq s \leq N^2$ and $\mathbf{w}^\alpha \cdot \mathbf{w}^\beta = \delta^{\alpha\beta}$. So, if $\mathbf{A}^\alpha$ is the folding of $\mathbf{w}^\alpha$ for all $1 \leq \alpha \leq s$ (i.e. $A^\alpha_{ij} = w^\alpha_{(j-1)N+i}$ for all $1 \leq i, j \leq N$), it follows that $\mathbf{A}^\alpha : \mathbf{A}^\beta = \delta^{\alpha\beta}$, each $\mathbf{A}^\alpha$ is skew-symmetric, and

$$\boldsymbol{\zeta} = \sum_{\alpha=1}^{s} \mu_\alpha \, \mathbf{A}^\alpha \otimes \mathbf{A}^\alpha.$$

Proposition 3.1 is constructive and could be used as a recipe for computing the tensors $\boldsymbol{\xi}, \boldsymbol{\zeta}$ which are necessary for the metriplectic ROM (9). On the other hand, an even simpler description of these objects can be found by appealing to some basic facts from exterior algebra (unfamiliar readers can find everything necessary in e.g. [30, Chapter 1] or [31, Chapter 19]). Beginning with the computation of $\boldsymbol{\xi}$, recall that any SS matrix (more formally, any antisymmetric $(0,2)$-tensor) decomposes as a sum of basis bivectors $\mathbf{e}_i \wedge \mathbf{e}_j = \mathbf{e}_i \otimes \mathbf{e}_j - \mathbf{e}_j \otimes \mathbf{e}_i$. In view of this, it is convenient to identify the matrix $\mathbf{L}$ with the bivector sum $\mathbf{L} = \sum_{j<i} L^{ij} \mathbf{e}_i \wedge \mathbf{e}_j$, where the functions $L^{ij}(\mathbf{x})$ may depend on the state $\mathbf{x}$. The action of $\mathbf{L}$ on a vector $\mathbf{v} \in \mathbb{R}^N$ is then identical to the matrix-vector product, since

$$\mathbf{L}\mathbf{v} = \sum_{j<i} \left(L^{ij} \left(\mathbf{e}_j \cdot \mathbf{v} \right) \mathbf{e}_i - L^{ij} \left(\mathbf{e}_i \cdot \mathbf{v} \right) \mathbf{e}_j \right) = \sum_{j<i} L^{ij} v_j \mathbf{e}_i - \sum_{j>i} L^{ji} v_j \mathbf{e}_i = \sum_{i,j} L^{ij} v_j \mathbf{e}_i,$$

where the skew-symmetry of $\mathbf{L}$ was used along with the relationship

$$(\mathbf{b} \wedge \mathbf{c}) \cdot \mathbf{a} = (-1)^{1(2+1)} \mathbf{a} \cdot (\mathbf{b} \wedge \mathbf{c}) = (\mathbf{a} \cdot \mathbf{c}) \mathbf{b} - (\mathbf{a} \cdot \mathbf{b}) \mathbf{c}, \qquad \mathbf{a}, \mathbf{b}, \mathbf{c} \in \mathbb{R}^N.$$

The advantage of this identification is a simple and coordinate-free solution to (6) which is computable following Proposition 3.1 and does not require the storage of a rank-3 tensor. In particular, choose an index k_1 such that $S^{k_1} \neq 0$ and define

$$\boldsymbol{\xi} = \mathbf{L} \wedge \mathbf{s}_{k_1}, \qquad \mathbf{s}_{k_1} = \frac{\mathbf{e}_{k_1}}{S^{k_1}}.$$

It follows from the definitions of $\mathbf{L}, \mathbf{s}_{k_1}$ and properties of the exterior product that $\boldsymbol{\xi}$ is totally antisymmetric and satisfies the equality

$$\boldsymbol{\xi} \left(\nabla S \right) = \left(\nabla S \cdot \mathbf{L} \right) \wedge \mathbf{s}_{k_1} + \left(\mathbf{s}_{k_1} \cdot \nabla S \right) \mathbf{L} = -\mathbf{L} \nabla S \wedge \mathbf{s}_{k_1} + \mathbf{L} = \mathbf{L},$$

so that $\boldsymbol{\xi}$ also solves (6). Therefore, the reduced tensor $\hat{\boldsymbol{\xi}}$ which solves the normal equations (8) is easily constructed through

$$\hat{\boldsymbol{\xi}} = \mathbf{U}^\intercal \boldsymbol{\xi} \left(\mathbf{U} \right) \mathbf{U} = \mathbf{U}^\intercal \mathbf{L} \mathbf{U} \wedge \mathbf{U}^\intercal \mathbf{s}_{k_1} = \hat{\mathbf{L}} \wedge \hat{\mathbf{s}}_{k_1},$$

where $\hat{\mathbf{L}} = \mathbf{U}^\intercal \mathbf{L} \mathbf{U}$ as in (5) and $\hat{\mathbf{s}}_{k_1} = \mathbf{U}^\intercal \mathbf{s}_{k_1}$. This implies that the structure-preserving counterpart to $\mathbf{L}$ (no longer called $\hat{\mathbf{L}}$!) is given in bivector form by

$$\hat{\boldsymbol{\xi}} \left(\nabla \hat{S} \right) = -\hat{\mathbf{L}} \nabla \hat{S} \wedge \hat{\mathbf{s}}_{k_1} + \left(\hat{\mathbf{s}}_{k_1} \cdot \nabla \hat{S} \right) \hat{\mathbf{L}}.$$

The degeneracy condition $\hat{\boldsymbol{\xi}} \left(\nabla \hat{S} \right) \nabla \hat{S} = 0$ is now satisfied by construction, since $\mathbf{L}$ is skew-symmetric and so

$$\hat{\boldsymbol{\xi}} \left(\nabla \hat{S} \right) \nabla \hat{S} = (0) \hat{\mathbf{s}}_{k_1} - \left(\hat{\mathbf{s}}_{k_1} \cdot \nabla \hat{S} \right) \hat{\mathbf{L}} \nabla \hat{S} + \left(\hat{\mathbf{s}}_{k_1} \cdot \nabla \hat{S} \right) \hat{\mathbf{L}} \nabla \hat{S} = \mathbf{0}.$$

Remark 3.2. As in the computation above, it is usually beneficial to avoid expressing multivectors in a traditional tensor basis, since the number of terms grows factorially with the degree. While bivectors are easily expressed coordinate-wise using the relations $\mathbf{e}_i \wedge \mathbf{e}_j = \mathbf{e}_i \otimes \mathbf{e}_j - \mathbf{e}_j \otimes \mathbf{e}_i$, basis trivectors already require a 6-term sum indexed over the permutation group S_3, producing many more intermediates which must be computed and stored.

It is similarly beneficial to rewrite the 4-tensor $\boldsymbol{\zeta}$. Note that the expression for $\boldsymbol{\zeta}$ in Proposition 3.1 is already somewhat useful for this purpose because it enables the computation of the metriplectic ROM (9) without actually storing or manipulating the full object. This is because the term entering (1) rewrites as

$$\mathbf{M} \nabla S = \boldsymbol{\zeta} \left(\nabla E, \nabla E \right) \nabla S = \sum_{\alpha=1}^{r} \lambda_\alpha \left(\nabla S \cdot \mathbf{A}^\alpha \nabla E \right) \mathbf{A}^\alpha \nabla E,$$

so that $\boldsymbol{\zeta}$ is accessed exclusively through the matrices $\mathbf{A}^\alpha$. However, following the construction in Proposition 3.1 still requires storing an array of skew-symmetric matrices which are extraordinarily sparse and often depend on the state $\mathbf{x}$, adding unnecessary computational expense. While the $\mathbf{x}$-dependence is not easy to handle, the storage and computation of the $\mathbf{A}^\alpha$ can be reduced with an exterior algebraic factorization as before.

Remark 3.3. In practice, it is usually easier to work with $\mathbf{m}^\alpha \leftarrow \sqrt{\lambda_\alpha} \mathbf{m}^\alpha$. From now on, we write $\mathbf{M} = \sum_\alpha \mathbf{m}^\alpha \otimes \mathbf{m}^\alpha$ with the understanding that the eigenvectors $\mathbf{m}^\alpha$ no longer have unit magnitude. The reader can check that the conclusions of Proposition 3.1 are unaffected by this change.

Again, it is useful to identify the matrices $\mathbf{A}^\alpha$ with the bivectors $\mathbf{A}^\alpha = \sum_{j<i} A^{\alpha,ij}\mathbf{e}_i \wedge \mathbf{e}_j$, so that for each $1 \leq \alpha \leq r$ there is a decomposition following Proposition 3.1,

$$\mathbf{A}^\alpha = \mathbf{a}^\alpha_{k_0} \wedge \mathbf{e}_{k_0}, \qquad \mathbf{a}^\alpha_{k_0} = \frac{\mathbf{m}^\alpha}{E^{k_0}},$$

where k_0 is an index such that $E^{k_0} \neq 0$. Since the normal equations (8) for $\hat{\boldsymbol{\zeta}}$ are solved when $\hat{\mathbf{A}}^\alpha = \mathbf{U}^\intercal \mathbf{A}^\alpha \mathbf{U}$, this means the reduced-order $\hat{\mathbf{A}}^\alpha$ are given by

$$\hat{\mathbf{A}}^\alpha = \hat{\mathbf{a}}^\alpha_{k_0} \wedge \mathbf{U}^{k_0}, \qquad \hat{\mathbf{a}}^\alpha_{k_0} = \frac{\mathbf{U}^\intercal \mathbf{m}^\alpha}{E^{k_0}},$$

where $\mathbf{U}^{k_0} = \mathbf{U}^\intercal \mathbf{e}_{k_0}$ denotes the k_0^{th} row of $\mathbf{U}$. This affords a simple representation for the matrix-vector products

$$\hat{\mathbf{A}}^\alpha \nabla \hat{E} = -\nabla \hat{E} \cdot \left(\hat{\mathbf{a}}^\alpha_{k_0} \wedge \mathbf{U}^{k_0}\right) = \left(\nabla \hat{E} \cdot \mathbf{U}^{k_0}\right) \hat{\mathbf{a}}^\alpha_{k_0} - \left(\nabla \hat{E} \cdot \hat{\mathbf{a}}^\alpha_{k_0}\right) \mathbf{U}^{k_0},$$

leading to greater efficiency in the numerical implementation. Moreover, it is easy to see that $\hat{\mathbf{A}}^\alpha$ remains skew-symmetric for all α, so that $\hat{\boldsymbol{\zeta}}\left(\nabla \hat{E}, \nabla \hat{E}\right) \nabla \hat{E} = \mathbf{0}$ by construction. Putting all of the computations in this Section together yields the following result which facilitates the numerical experiments in Section 5.

Theorem 3.4 (Metriplectic structure-preserving ROM)**.** *Suppose* $\mathbf{L}$, $\mathbf{M} = \sum_\alpha \mathbf{m}^\alpha \otimes \mathbf{m}^\alpha$, E *and* S *describe a metriplectic dynamical system* (1) *with state* $\mathbf{x} \in \mathbb{R}^N$ *and associated tensors* $\boldsymbol{\xi}, \boldsymbol{\zeta}$ *satisfying* (6) *and* (7) *for all* $\mathbf{x}$*. Suppose* $\hat{\mathbf{x}} \in \mathbb{R}^n$ *is a low-dimensional approximation to the state in the sense that* $\mathbf{x} \approx \tilde{\mathbf{x}} = \mathbf{x}_0 + \mathbf{U}\hat{\mathbf{x}}$ *for some initial state* $\mathbf{x}_0 \in \mathbb{R}^N$ *and linear mapping* $\mathbf{U} \in \mathbb{R}^{N \times n}$*. Let* $1 \leq k_0, k_1 \leq N$ *be indices such that* $E^{k_0}(\mathbf{x}) \neq 0$ *and* $S^{k_1}(\mathbf{x}) \neq 0$ *for all* $\mathbf{x}$ *(otherwise the conclusion holds locally) and define*

$$\hat{\boldsymbol{\xi}} = \hat{\mathbf{L}} \wedge \hat{\mathbf{s}}_{k_1}, \qquad \hat{\mathbf{s}}_{k_1} = \frac{\mathbf{U}^{k_1}}{S^{k_1}},$$
$$\hat{\mathbf{A}}^\alpha = \hat{\mathbf{a}}^\alpha_{k_0} \wedge \mathbf{U}^{k_0}, \qquad \hat{\mathbf{a}}^\alpha_{k_0} = \frac{\mathbf{U}^\intercal \mathbf{m}^\alpha}{E^{k_0}},$$

where $\hat{\mathbf{L}} = \mathbf{U}^\intercal \mathbf{L} \mathbf{U}$ *and* $\hat{\boldsymbol{\zeta}} = \sum_\alpha \hat{\mathbf{A}}^\alpha \otimes \hat{\mathbf{A}}^\alpha$*. Then, denoting* $\hat{F} = F \circ \tilde{\mathbf{x}}$ *for any function* F*, the ROM*

$$\begin{aligned}\dot{\hat{\mathbf{x}}} &= \left\{\hat{\mathbf{x}}, \nabla \hat{E}\left(\hat{\mathbf{x}}\right)\right\} + \left[\hat{\mathbf{x}}, \nabla \hat{S}\left(\hat{\mathbf{x}}\right)\right] \\ &:= \hat{\boldsymbol{\xi}}\left(\tilde{\mathbf{x}}\right)\left(\nabla \hat{S}\left(\hat{\mathbf{x}}\right)\right) \nabla \hat{E}\left(\hat{\mathbf{x}}\right) + \hat{\boldsymbol{\zeta}}\left(\tilde{\mathbf{x}}\right)\left(\nabla \hat{E}\left(\hat{\mathbf{x}}\right), \nabla \hat{E}\left(\hat{\mathbf{x}}\right)\right) \nabla \hat{S}\left(\hat{\mathbf{x}}\right),\end{aligned}$$

preserves the original metriplectic structure. Suppressing the potential state dependence, this has the explicit representation

$$\begin{aligned}\dot{\hat{\mathbf{x}}} = &\left(\nabla \hat{E} \cdot \hat{\mathbf{L}} \nabla \hat{S}\right) \hat{\mathbf{s}}_{k_1} - \left(\hat{\mathbf{s}}_{k_1} \cdot \nabla \hat{E}\right) \hat{\mathbf{L}} \nabla \hat{S} + \left(\hat{\mathbf{s}}_{k_1} \cdot \nabla \hat{S}\right) \hat{\mathbf{L}} \nabla \hat{E} \\ &+ \sum_{\alpha=1}^{r} \left(\mathbf{A}^\alpha \nabla \hat{E} \cdot \nabla \hat{S}\right) \mathbf{A}^\alpha \nabla \hat{E},\end{aligned}$$

where $\mathbf{A}^\alpha \nabla \hat{E}$ *is expressed in terms of* $\hat{\mathbf{a}}_{k_0}$ *and the* k_0^{th} *row of* $\mathbf{U}$ *as*

$$\hat{\mathbf{A}}^\alpha \nabla \hat{E} = \left(\nabla \hat{E} \cdot \mathbf{U}^{k_0}\right) \hat{\mathbf{a}}^\alpha_{k_0} - \left(\nabla \hat{E} \cdot \hat{\mathbf{a}}^\alpha_{k_0}\right) \mathbf{U}^{k_0}.$$

Remark 3.5. Note that the quantities $\hat{\mathbf{L}}, \mathbf{A}^\alpha, \mathbf{m}^\alpha, \hat{\mathbf{s}}_{k_1}, \hat{\mathbf{a}}^\alpha_{k_0}$ may depend on $\tilde{\mathbf{x}}$, so that the metriplectic ROM generally depends on the full N-dimensional input space. On the other hand, there are many special cases where this dependence can be mitigated or removed entirely. Example 5.2 in Section 5 gives one such illustration. Future work will also consider hyper-reduction strategies such as DEIM [32] for this purpose.

4. Error estimate

It is important to know that the metriplectic ROM described in Theorem 3.4 will converge to the true solution as the reduced dimension n approaches the full resolution N. The results of this Section show that this is indeed the case when the mapping U is generated by POD and the snapshot matrix is augmented with gradient information. To explain why, let $\|\cdot\|$ denote the usual norm in $L^2([0,T])$ and denote $\mathbf{w} = \mathbf{x} - \mathbf{x}_0$.

Choose weighting parameters $\mu, \nu \in \mathbb{R}$ as well as n_t instants $t_i \in [0, T]$ and consider the snapshot matrix $\mathbf{Y} \in \mathbb{R}^{N \times 3n_t}$,

$$\mathbf{Y} = \begin{pmatrix} \mathbf{w}(t_i) & \mu \nabla E(\mathbf{x}(t_i)) & \nu \nabla S(\mathbf{x}(t_i)) \end{pmatrix}_{i=1}^{n_t}.$$

Remark 4.1. For simplicity, all experiments in Section 5 use the weights $\mu = \nu = 1$.

Suppose $\mathbf{Y}$ has rank r and denote by σ_i the i^{th} singular value of Y. Then, if $\mathbf{U} \in \mathbb{R}^{N \times n}$ is the rank $n < r$ matrix of left singular vectors and $\mathbf{P}^\perp = \mathbf{I} - \mathbf{U}\mathbf{U}^\intercal$ is orthogonal projection onto the complement of Span $(\mathbf{U})$, standard POD theory (see e.g. [29]) implies that

$$\left\|\mathbf{P}^\perp \mathbf{w}\right\|^2 + \mu^2 \left\|\mathbf{P}^\perp \nabla E\right\|^2 + \nu^2 \left\|\mathbf{P}^\perp \nabla S\right\|^2 = \sum_{j>n} \sigma_j^2. \tag{10}$$

Recall also the Lipschitz constant and logarithmic Lipschitz constant, denoted for a general function F between metric spaces as

$$C_F = \sup_{\mathbf{u} \neq \mathbf{v}} \frac{|F(\mathbf{u}) - F(\mathbf{v})|}{|\mathbf{u} - \mathbf{v}|}, \qquad c_F = \sup_{\mathbf{u} \neq \mathbf{v}} \frac{(\mathbf{u} - \mathbf{v}) \cdot (F(\mathbf{u}) - F(\mathbf{v}))}{|\mathbf{u} - \mathbf{v}|^2}.$$

This can be used to derive the following result for the present scheme.

Theorem 4.2. *Let $\mathbf{x}(t)$ denote the solution to the FOM (1) with initial condition $\mathbf{x}_0$ and let $\hat{\mathbf{x}}(t)$ denote the solution to the ROM (9) with $\hat{\boldsymbol{\xi}} = \mathbf{U}^\intercal \boldsymbol{\xi}(\mathbf{U})\,\mathbf{U}$, $\hat{\boldsymbol{\zeta}} = \mathbf{U}^\intercal \boldsymbol{\zeta}(\mathbf{U}, \mathbf{U})\,\mathbf{U}$ and initial condition $\hat{\mathbf{x}}_0 = \mathbf{0}$. Suppose $\boldsymbol{\xi}, \boldsymbol{\zeta}, \nabla E, \nabla S$ are Lipschitz continuous in space and L^2-integrable in time. Then, the approximation error satisfies*

$$\|\mathbf{x} - (\mathbf{x}_0 + \mathbf{U}\hat{\mathbf{x}})\|^2 \leq C(T, \boldsymbol{\xi}, \boldsymbol{\zeta}, \mu, \nu) \sum_{j>r} \sigma_j^2.$$

Proof. First, suppose $\mathbf{x}_0 = \mathbf{0}$, so that $\tilde{\mathbf{x}} = \mathbf{U}\hat{\mathbf{x}}$ and the approximation error becomes

$$\mathbf{x} - \mathbf{U}\hat{\mathbf{x}} = (\mathbf{x} - \mathbf{U}\mathbf{U}^\intercal \mathbf{x}) + (\mathbf{U}\mathbf{U}^\intercal \mathbf{x} - \mathbf{U}\hat{\mathbf{x}}) = \mathbf{P}^\perp \mathbf{x} + \mathbf{y}.$$

For cleanliness of notation, we denote $\bar{\mathbf{X}} = \mathbf{U}\mathbf{U}^\intercal \mathbf{X}$ and

$$\mathbf{F}(\mathbf{x}) = \bar{\boldsymbol{\xi}}(\mathbf{x}) \left(\bar{\nabla} S(\mathbf{x})\right) \bar{\nabla} E(\mathbf{x}) + \bar{\boldsymbol{\zeta}}(\mathbf{x}) \left(\bar{\nabla} E(\mathbf{x}), \bar{\nabla} E(\mathbf{x})\right) \bar{\nabla} S(\mathbf{x}).$$

It follows from (1) and (9) that the error decomposes into three terms,

$$\begin{aligned}
\dot{\mathbf{y}} = \dot{\tilde{\mathbf{x}}} - \mathbf{U}\dot{\hat{\mathbf{x}}} &= \bar{\boldsymbol{\xi}}(\nabla S(\mathbf{x}))\,\nabla E(\mathbf{x}) + \bar{\boldsymbol{\zeta}}(\nabla E(\mathbf{x}), \nabla E(\mathbf{x}))\,\nabla S(\mathbf{x}) - \mathbf{F}(\mathbf{U}\hat{\mathbf{x}}) \\
&= \bar{\boldsymbol{\xi}}(\nabla S(\mathbf{x}))\,\nabla E(\mathbf{x}) + \bar{\boldsymbol{\zeta}}(\nabla E(\mathbf{x}), \nabla E(\mathbf{x}))\,\nabla S(\mathbf{x}) - \mathbf{F}(\mathbf{x}) \\
&\quad + (\mathbf{F}(\mathbf{x}) - \mathbf{F}(\bar{\mathbf{x}})) + (\mathbf{F}(\bar{\mathbf{x}}) - \mathbf{F}(\mathbf{U}\hat{\mathbf{x}})) \\
&:= \mathbf{T}_1 + \mathbf{T}_2 + \mathbf{T}_3.
\end{aligned}$$

Suppressing the $\mathbf{x}$ argument, $\mathbf{T}_1$ can be written as

$$\begin{aligned}
\mathbf{T}_1 = \bar{\boldsymbol{\xi}}\left(\mathbf{P}^\perp \nabla S\right) \nabla E + \bar{\boldsymbol{\xi}}\left(\bar{\nabla} S\right) \mathbf{P}^\perp \nabla E \\
+ \bar{\boldsymbol{\zeta}}\left(\mathbf{P}^\perp \nabla E, \nabla E\right) \nabla S + \bar{\boldsymbol{\zeta}}\left(\bar{\nabla} E, \mathbf{P}^\perp \nabla E\right) \bar{\nabla} S + \bar{\boldsymbol{\zeta}}\left(\bar{\nabla} E, \bar{\nabla} E\right) \mathbf{P}^\perp \nabla S,
\end{aligned}$$

so that its norm can be estimated as

$$\begin{aligned}
|\mathbf{T}_1| &\leq \left(\left|\bar{\boldsymbol{\xi}}\right| \left|\bar{\nabla} S\right| + \left|\bar{\boldsymbol{\zeta}}\right| |\nabla E|\,|\nabla S| + \left|\bar{\boldsymbol{\zeta}}\right| \left|\bar{\nabla} E\right| \left|\bar{\nabla} S\right|\right) \left|\mathbf{P}^\perp \nabla E\right| \\
&\quad + \left(\left|\bar{\boldsymbol{\xi}}\right| |\nabla E| + \left|\bar{\boldsymbol{\zeta}}\right| \left|\bar{\nabla} E\right|^2\right) \left|\mathbf{P}^\perp \nabla S\right| \\
&:= f_1 \left|\mathbf{P}^\perp \nabla E\right| + f_2 \left|\mathbf{P}^\perp \nabla S\right|.
\end{aligned}$$

The term $\mathbf{T}_2$ can be rewritten in a similar fashion. Collecting all terms involving $\boldsymbol{\xi}$ yields

$$\begin{aligned}
&\bar{\boldsymbol{\xi}}(\mathbf{x}) \left(\bar{\nabla} S(\mathbf{x})\right) \bar{\nabla} E(\mathbf{x}) - \bar{\boldsymbol{\xi}}(\bar{\mathbf{x}}) \left(\bar{\nabla} S(\bar{\mathbf{x}})\right) \bar{\nabla} E(\bar{\mathbf{x}}) \\
&\quad = \left(\bar{\boldsymbol{\xi}}(\mathbf{x}) - \bar{\boldsymbol{\xi}}(\bar{\mathbf{x}})\right) \left(\bar{\nabla} S(\mathbf{x})\right) \bar{\nabla} E(\mathbf{x}) \\
&\qquad + \bar{\boldsymbol{\xi}}(\bar{\mathbf{x}}) \left(\bar{\nabla} S(\mathbf{x}) - \bar{\nabla} S(\bar{\mathbf{x}})\right) \bar{\nabla} E(\mathbf{x}) \\
&\qquad + \bar{\boldsymbol{\xi}}(\bar{\mathbf{x}}) \left(\bar{\nabla} S(\bar{\mathbf{x}})\right) \left(\bar{\nabla} E(\mathbf{x}) - \bar{\nabla} E(\bar{\mathbf{x}})\right),
\end{aligned}$$

while collecting the terms involving $\boldsymbol{\zeta}$ gives

$$\begin{aligned}
&\bar{\boldsymbol{\zeta}}(\mathbf{x})\left(\bar{\nabla}E(\mathbf{x}),\bar{\nabla}E(\mathbf{x})\right)\bar{\nabla}S(\mathbf{x})-\bar{\boldsymbol{\zeta}}\left(\bar{\mathbf{x}}\right)\left(\bar{\nabla}E\left(\bar{\mathbf{x}}\right),\bar{\nabla}E\left(\bar{\mathbf{x}}\right)\right)\bar{\nabla}S\left(\bar{\mathbf{x}}\right)\\
&\quad=\left(\bar{\boldsymbol{\zeta}}(\mathbf{x})-\bar{\boldsymbol{\zeta}}\left(\bar{\mathbf{x}}\right)\right)\left(\bar{\nabla}E(\mathbf{x}),\bar{\nabla}E(\mathbf{x})\right)\bar{\nabla}S(\mathbf{x})\\
&\qquad+\bar{\boldsymbol{\zeta}}\left(\bar{\mathbf{x}}\right)\left(\bar{\nabla}E(\mathbf{x})-\bar{\nabla}E\left(\bar{\mathbf{x}}\right),\bar{\nabla}E(\mathbf{x})\right)\bar{\nabla}S(\mathbf{x})\\
&\qquad+\bar{\boldsymbol{\zeta}}\left(\bar{\mathbf{x}}\right)\left(\bar{\nabla}E\left(\bar{\mathbf{x}}\right),\bar{\nabla}E(\mathbf{x})-\bar{\nabla}E\left(\bar{\mathbf{x}}\right)\right)\bar{\nabla}S(\mathbf{x})\\
&\qquad+\bar{\boldsymbol{\zeta}}\left(\bar{\mathbf{x}}\right)\left(\bar{\nabla}E\left(\bar{\mathbf{x}}\right),\bar{\nabla}E\left(\bar{\mathbf{x}}\right)\right)\left(\bar{\nabla}S(\mathbf{x})-\bar{\nabla}S\left(\bar{\mathbf{x}}\right)\right).
\end{aligned}$$

Hence, the norm of $|\mathbf{T}_2|$ is bounded above by

$$\begin{aligned}
|\mathbf{T}_2|&\leq|\mathbf{U}\mathbf{U}^{\intercal}|\left|\bar{\nabla}S(\mathbf{x})\right|\left|\bar{\nabla}E(\mathbf{x})\right|\left(C_{\xi}+C_{\zeta}\left|\bar{\nabla}E(\mathbf{x})\right|\right)\left|\mathbf{P}^{\perp}\mathbf{x}\right|\\
&\quad+|\mathbf{U}\mathbf{U}^{\intercal}|\left|\bar{\boldsymbol{\xi}}\left(\bar{\mathbf{x}}\right)\right|\left(C_{\nabla S}\left|\bar{\nabla}E(\mathbf{x})\right|+C_{\nabla E}\left|\bar{\nabla}S\left(\bar{\mathbf{x}}\right)\right|\right)\left|\mathbf{P}^{\perp}\mathbf{x}\right|\\
&\quad+C_{\nabla E}|\mathbf{U}\mathbf{U}^{\intercal}|\left|\bar{\boldsymbol{\zeta}}\left(\bar{\mathbf{x}}\right)\right|\left|\bar{\nabla}S(\mathbf{x})\right|\left(\left|\bar{\nabla}E(\mathbf{x})\right|+\left|\bar{\nabla}E\left(\bar{\mathbf{x}}\right)\right|\right)\left|\mathbf{P}^{\perp}\mathbf{x}\right|\\
&\quad+C_{\nabla S}|\mathbf{U}\mathbf{U}^{\intercal}|\left|\bar{\boldsymbol{\zeta}}\left(\bar{\mathbf{x}}\right)\right|\left|\bar{\nabla}E\left(\bar{\mathbf{x}}\right)\right|^{2}\left|\mathbf{P}^{\perp}\mathbf{x}\right|\\
&:=f_3\left|\mathbf{P}^{\perp}\mathbf{x}\right|
\end{aligned}$$

The assumptions of the Theorem imply that the supremum of f_3 over $\mathbf{x}\neq\bar{\mathbf{x}}$ is finite, so the above inequality combined with $\mathbf{T}_2=\mathbf{F}\left(\mathbf{x}\right)-\mathbf{F}\left(\bar{\mathbf{x}}\right)$ implies that $C_F\leq\sup f_3<\infty$ and the Lipschitz constant C_F exists. As $|c_F|\leq C_F$, it follows that the logarithmic Lipschitz constant c_F exists also.

Therefore, testing each term of $\dot{\mathbf{y}}$ against $\mathbf{y}$ and estimating with Cauchy-Schwarz yields

$$\begin{aligned}
\mathbf{y}\cdot\mathbf{T}_1&\leq\left(f_1\left|\mathbf{P}^{\perp}\nabla E\right|+f_2\left|\mathbf{P}^{\perp}\nabla S\right|\right)|\mathbf{y}|,\\
\mathbf{y}\cdot\mathbf{T}_2&\leq f_3\left|\mathbf{P}^{\perp}\mathbf{x}\right||\mathbf{y}|,\\
\mathbf{y}\cdot\mathbf{T}_3&=\frac{\mathbf{y}\cdot\mathbf{T}_3}{|\mathbf{y}|^2}|\mathbf{y}|^2\leq c_F|\mathbf{y}|^2,
\end{aligned}$$

where the scalar functions f_j may depend on t since their terms may depend on $\mathbf{x}$. Noting that $|\dot{\mathbf{y}}|=(1/|\mathbf{y}|)(\mathbf{y}\cdot\dot{\mathbf{y}})$ and using the estimates above yields the inequality

$$|\dot{\mathbf{y}}|-c_F|\mathbf{y}|\leq f_1\left|\mathbf{P}^{\perp}\nabla E\right|+f_2\left|\mathbf{P}^{\perp}\nabla S\right|+f_3\left|\mathbf{P}^{\perp}\mathbf{x}\right|.$$

Multiplying both sides by the integrating factor e^{-c_Ft} and applying Gronwall's inequality for $t\in[0,T]$ then gives

$$|\mathbf{y}|\leq\int_0^t e^{c_F(t-\tau)}\left(f_1\left|\mathbf{P}^{\perp}\nabla E\right|+f_2\left|\mathbf{P}^{\perp}\nabla S\right|+f_3\left|\mathbf{P}^{\perp}\mathbf{x}\right|\right)d\tau,$$

where we have used that $\mathbf{y}(0)=\mathbf{0}$. Then, Cauchy-Schwarz along with $T\geq t$ imply

$$\begin{aligned}
|\mathbf{y}|^2&\leq C_4\left\|f_1\left|\mathbf{P}^{\perp}\nabla E\right|+f_2\left|\mathbf{P}^{\perp}\nabla S\right|+f_3\left|\mathbf{P}^{\perp}\mathbf{x}\right|\right\|^2\\
&\leq C_4\left(\|f_1\|^2\left\|\mathbf{P}^{\perp}\nabla E\right\|^2+\|f_2\|^2\left\|\mathbf{P}^{\perp}\nabla S\right\|^2+\|f_3\|^2\left\|\mathbf{P}^{\perp}\mathbf{x}\right\|^2\right),
\end{aligned}$$

where $C_4=\int_0^T e^{2c_F(T-\tau)}=(e^{2c_FT}-1)/2c_F$. Finally, integrating both sides in t yields

$$\|\mathbf{y}\|^2\leq TC_4\left(\|f_1\|^2\left\|\mathbf{P}^{\perp}\nabla E\right\|^2+\|f_2\|^2\left\|\mathbf{P}^{\perp}\nabla S\right\|^2+\|f_3\|^2\left\|\mathbf{P}^{\perp}\mathbf{x}\right\|^2\right),$$

so that in view of (10) the error can be estimated as

$$\begin{aligned}
\|\mathbf{x}-\mathbf{U}\hat{\mathbf{x}}\|^2&\leq\left\|\mathbf{P}^{\perp}\mathbf{x}\right\|^2+\|\mathbf{y}\|^2\\
&\leq\left(1+TC_4\left(\frac{\|f_1\|^2}{\mu^2}+\frac{\|f_2\|^2}{\nu^2}+\|f_3\|^2\right)\right)\sum_{j>r}\sigma_j^2=C\sum_{j>r}\sigma_j^2,
\end{aligned}$$

as desired. To complete the argument, consider what happens when $\mathbf{x}_0\neq\mathbf{0}$. In this case, $\mathbf{w}=\mathbf{x}-\mathbf{x}_0$ satisfies $\dot{\mathbf{w}}=\dot{\mathbf{x}}$ and the ROM error is

$$\mathbf{x}-(\mathbf{x}_0+\mathbf{U}\hat{\mathbf{x}})=(\mathbf{w}-\bar{\mathbf{w}})+(\bar{\mathbf{w}}-\mathbf{U}\hat{\mathbf{x}})=\mathbf{P}^{\perp}\mathbf{w}+\mathbf{z}.$$

Since $\left\|\mathbf{P}^{\perp}\mathbf{w}\right\|$ is controlled by (10) and $\mathbf{z}$ rewrites as

$$\begin{aligned}\dot{\mathbf{z}} &= \dot{\hat{\mathbf{x}}} - \mathbf{F}\left(\mathbf{x}_0 + \mathbf{U}\hat{\mathbf{x}}\right)\\ &= \left(\dot{\hat{\mathbf{x}}} - \mathbf{F}(\mathbf{x})\right) + \left(\mathbf{F}(\mathbf{x}) - \bar{\mathbf{F}}\left(\mathbf{x}_0 + \bar{\mathbf{w}}\right)\right) + \left(\mathbf{F}\left(\mathbf{x}_0 + \bar{\mathbf{w}}\right) - \mathbf{F}\left(\mathbf{x}_0 + \mathbf{U}\hat{\mathbf{x}}\right)\right),\end{aligned}$$

this case can be analyzed identically to the case $\mathbf{x}_0 = \mathbf{0}$. □

5. Examples

This section evaluates the performance of the metriplectic ROM (referred to as the SP-ROM) proposed in Theorem 3.4 on some benchmark problems. To create a fair comparison, its performance is measured against that of the standard Galerkin POD-ROM (3) referred to as the G-ROM and a mild extension (5) of the Hamiltonian POD-ROM from [16] referred to as the EH-ROM which approximately preserves the compatibility conditions (2). The error metrics used for this purpose are the relative ℓ^2 error and the maximum ℓ^2 error, denoted respectively as

$$\mathcal{E}_r\left(\tilde{\mathbf{x}}\right) = \sqrt{\frac{\sum_i \left|\mathbf{x}(t_i) - \tilde{\mathbf{x}}(t_i)\right|^2}{\sum_i \left|\mathbf{x}(t_i)\right|^2}}, \qquad \mathcal{E}_\infty\left(\tilde{\mathbf{x}}\right) = \max_i \left|\mathbf{x}(t_i) - \tilde{\mathbf{x}}(t_i)\right|,$$

where $\mathbf{x}$ is the true solution and $1 \le i \le n_t$ are the indices of the discretization points $t_i \in [0, T]$. Besides the error metrics, the energy difference $|E(T) - E_0|$ ($E_0 = E(t_0)$) is reported as well as the online computational time in seconds necessary for integrating each model. This collection of data provides a rough measure of model quality which is used to draw conclusions about ROM performance. The experiments chosen are a low-dimensional example motivated by gas kinetics and an infinite-dimensional example coming from elasticity theory. In both cases, the initial conditions are parameterized and used to train the mapping $\mathbf{U}$, and the ROMs are used to predict unseen solutions given relevant initial data. In all cases, the SciPy [33] interface to LSODA [34] is used to integrate the resulting ODE systems with an error tolerance of 10^{-14}. All experiments are carried out on a 2022 M1 MacBook Pro with 32GB of RAM.

5.1. **Two gas containers exchanging heat and volume.** To verify the correctness of the claims in Theorem 4.2 and Theorem 3.4, it is useful to consider the following low-dimensional example from [35]. Two gas containers are allowed to exchange heat and volume on either side of a wall. The state variable is $\mathbf{x} = \begin{pmatrix} q & p & S_1 & S_2 \end{pmatrix}^{\intercal} \in \mathbb{R}^4$ where q, p are the position resp. momentum of the wall and S_1, S_2 are the entropy of the gases in the respective containers. The entropy function is then $S = S_1 + S_2$ and the energy function is

$$E(\mathbf{x}) = \frac{p^2}{2m} + E_1 + E_2 := \frac{p^2}{2m} + \left(\frac{e^{\frac{S_1}{(Nk_B)}}}{\hat{c}\, q}\right)^{\frac{2}{3}} + \left(\frac{e^{\frac{S_2}{(Nk_B)}}}{\hat{c}\,(2-q)}\right)^{\frac{2}{3}}, \quad \hat{c} = \left(\frac{4\pi m^2}{3h^2 N}\right)^{\frac{3}{2}} \frac{e^{\frac{5}{2}}}{N},$$

where m is the mass of the wall, N is the number of gas particles, h is the Planck constant and k_B is the Boltzmann constant. For the rest of the discussion, a normalization is assumed such that $M = Nk_B = \hat{c} = 1$. This system then has the metriplectic form

$$\begin{pmatrix} \dot{q} \\ \dot{p} \\ \dot{S}_1 \\ \dot{S}_2 \end{pmatrix} = \begin{pmatrix} p \\ \frac{2}{3}\left(\frac{E_1}{q} - \frac{E_2}{2-q}\right) \\ \frac{\gamma}{T_1}\left(\frac{1}{T_1} - \frac{1}{T_2}\right) \\ \frac{-\gamma}{T_2}\left(\frac{1}{T_1} - \frac{1}{T_2}\right) \end{pmatrix} = \mathbf{L}\nabla E(\mathbf{x}) + \mathbf{M}(\mathbf{x})\nabla S,$$

where $T_i = \partial_{S_i} E_i = (2/3)E_i$, $\nabla S = \begin{pmatrix} 0 & 0 & 1 & 1 \end{pmatrix}^{\top}$, γ regulates the degree of heat transfer, and

$$\mathbf{L} = \begin{pmatrix} 0 & 1 & 0 & 0 \\ -1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}, \quad \mathbf{M} = \gamma \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & T_1^{-2} & -(T_1T_2)^{-1} \\ 0 & 0 & -(T_1T_2)^{-1} & T_2^{-2}, \end{pmatrix}, \quad \nabla E = \frac{2}{3}\begin{pmatrix} -\left(\frac{e^{2S_1}}{p^5}\right)^{\frac{1}{3}} + \left(\frac{e^{2S_2}}{p^5}\right)^{\frac{1}{3}} \\ (3/2)\, q \\ E_1 \\ E_2 \end{pmatrix}.$$

Note that the matrix $\mathbf{M}$ decomposes (nonuniquely) as

$$\mathbf{M} = \mathbf{m} \otimes \mathbf{m}, \qquad \mathbf{m} = \sqrt{\gamma}\begin{pmatrix} 0 & 0 & T_1^{-1} & -T_2^{-1} \end{pmatrix}^{\intercal},$$

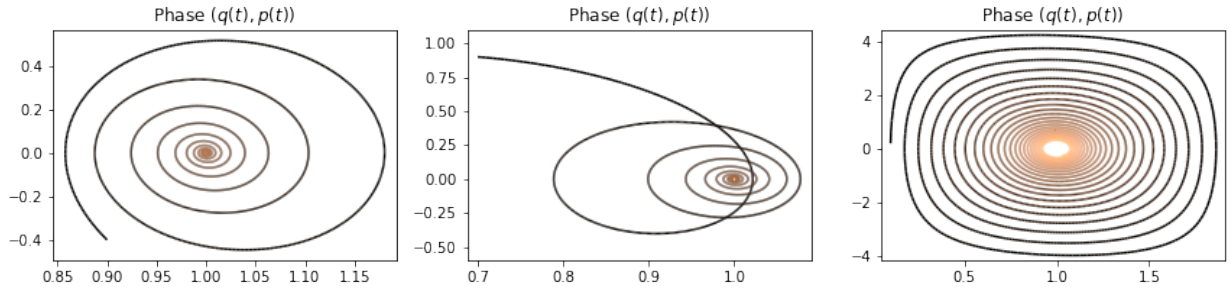

FIGURE 1. The phase portraits of three qualitatively different solutions to the gas container problem with corresponding initial conditions $\begin{pmatrix}0.9 & -0.4 & 2.4 & 2.0\end{pmatrix}^\intercal$ (left), $\begin{pmatrix}0.7 & 0.9 & 1.1 & 2.9\end{pmatrix}^\intercal$ (middle), and $\begin{pmatrix}0.1 & 0.2 & 1.6 & 1.8\end{pmatrix}^\intercal$ (right).

T	n	Method	$\mathcal{E}_r$ %	$\mathcal{E}_\infty$	$\lvert E(T)-E_0\rvert$	Time (s)	T	n	Method	$\mathcal{E}_r$ %	$\mathcal{E}_\infty$	$\lvert E(T)-E_0\rvert$	Time (s)
8	-	FOM	-	-	5.507×10^{-14}	0.07517	32	-	FOM	-	-	1.279×10^{-13}	0.2323
	2	SP-ROM	15.84	0.9166	1.066×10^{-13}	0.01457		2	SP-ROM	14.71	0.9166	1.243×10^{-13}	0.02071
		EH-ROM	16.00	0.9162	0.3127	0.01631			EH-ROM	15.29	0.9162	3.357×10^{-3}	0.02199
		G-ROM	58.96	2.436	46.29	0.02356			G-ROM	182.9	7.188	965.5	0.04956
	3	SP-ROM	7.462	0.2338	5.861×10^{-14}	0.07991		3	SP-ROM	7.367	0.2338	1.279×10^{-13}	0.2262
		EH-ROM	7.303	0.2246	0.4692	0.06545			EH-ROM	9.050	0.3677	1.378	0.1906
		G-ROM	6.808	0.2244	0.4364	0.06068			G-ROM	8.971	0.3754	1.422	0.1843
	4	SP-ROM	3.968×10^{-12}	3.465×10^{-13}	4.619×10^{-14}	0.07635		4	SP-ROM	1.005×10^{-11}	6.611×10^{-13}	2.025×10^{-13}	0.1734
		EH-ROM	5.252×10^{-12}	3.988×10^{-13}	1.385×10^{-13}	0.07155			EH-ROM	1.099×10^{-11}	8.299×10^{-13}	1.421×10^{-13}	0.1683
		G-ROM	6.769×10^{-12}	3.934×10^{-13}	2.345×10^{-13}	0.05992			G-ROM	1.062×10^{-11}	6.225×10^{-13}	2.984×10^{-13}	0.1725

TABLE 1. Results of the gas container experiment. Dashes indicate "not applicable" to the case of the FOM.

so that by choosing indices k_0, k_1 such that $E^{k_0} \neq 0$ and $S^{k_1} \neq 0$ the tensors $\boldsymbol{\zeta}, \boldsymbol{\xi}$ can be computed following Proposition 3.1. The choice $k_0 = k_1 = 3$ yields the reduced-order objects

$$\hat{\boldsymbol{\xi}} = \hat{\mathbf{L}} \wedge \mathbf{U}^3, \qquad \hat{\boldsymbol{\zeta}}(\tilde{\mathbf{x}}) = \hat{\mathbf{A}}(\tilde{\mathbf{x}}) \otimes \hat{\mathbf{A}}(\tilde{\mathbf{x}}), \qquad \hat{\mathbf{A}}(\tilde{\mathbf{x}}) = \frac{3}{2}\frac{\mathbf{U}^\intercal \mathbf{m}(\tilde{\mathbf{x}})}{E_1(\tilde{\mathbf{x}})} \wedge \mathbf{U}^3,$$

making it clear that $\hat{\boldsymbol{\xi}}$ can be precomputed while some components of $\hat{\boldsymbol{\zeta}}$ must be computed online. For the present experiment, ROM performance is compared at various levels of compression when integrated over two different lengths of time, one of which extends away from the training regime. The initial state $\mathbf{x}_0$ is assumed to lie in $[0.08, 1.8] \times [-1, 1] \times [1, 3] \times [1, 3]$, and snapshots from 25 simulations with $\gamma = 8$ and random, uniformly distributed initial data are used to train the POD approximation $\mathbf{U}$. Some representative solutions are displayed in Figure 1. Snapshots of the shifted solution $\mathbf{x} - \mathbf{x}_0$ as well as the gradient $\nabla E(\mathbf{x})$ are saved every 0.02 time instants in the interval $[0, 8]$ during training and concatenated to form the snapshot matrix $\mathbf{Y}$. As ∇S is a constant independent of $\mathbf{x}$, it is included only once as the final column of the snapshot matrix.

The performance of each ROM is tested using the solution with corresponding initial condition $\mathbf{x}_0 = \begin{pmatrix}0.9 & -0.4 & 2.4 & 2.0\end{pmatrix}^\intercal$ not included in the training set. Table 5.1 compares the errors that arise when integrating to $T = 8$ where snapshots end, as well as when integrating to $T = 32$ which extends well past this point. Figures 2 and 3 display the FOM and ROM positions q, entropies S, and energy variations $E - E_0$ over each time interval when $n = 2, 3, 4$. Notice that all methods converge with refinement as expected, but only the proposed SP-ROM is capable of preserving the initial energy (to order 10^{-13}) as well as producing a reasonable entropy profile regardless of the reduced dimension. It is also interesting to note that the simple G-ROM can produce quite unphysical results in terms of its energy and entropy profiles, including the rapid energy increase seen in Figure 3. This shows that there are clear benefits to preserving the metriplectic structure of the original system, especially in the presence of downstream procedures which rely on accurate estimates of such quantities.

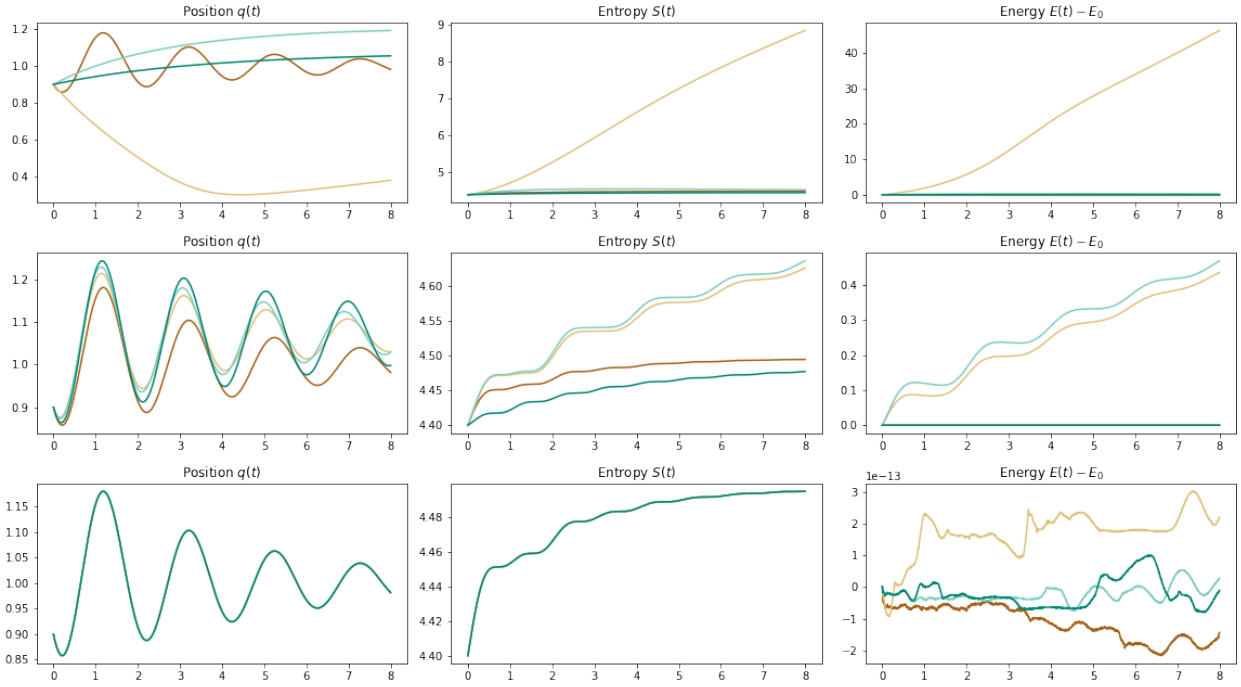

FIGURE 2. A comparison of ROM solutions for the 4-dimensional gas container example when $T = 8$ and $n = 2, 3, 4$, respectively. Plotted are the Exact Solution, G-ROM, EH-ROM, and SP-ROM. Observe the convergence as predicted in Theorem 4.2.

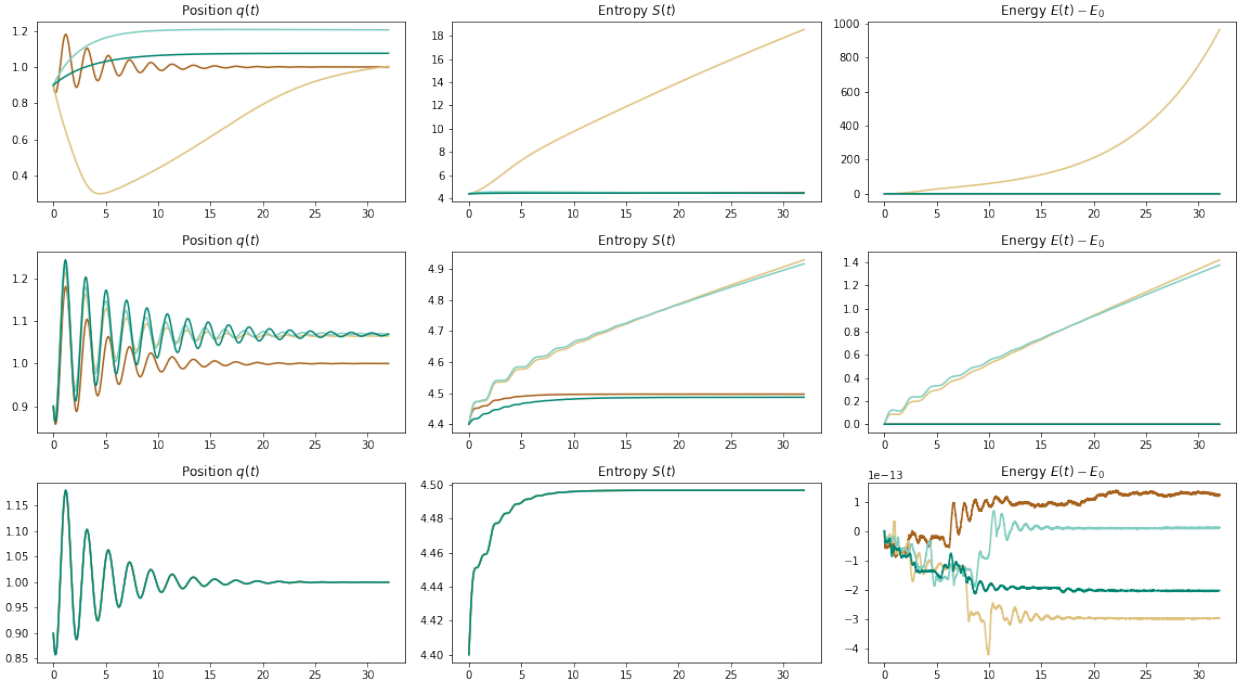

FIGURE 3. A comparison of ROM solutions for the 4-dimensional gas container example when $T = 32$ and $n = 2, 3, 4$, respectively. Plotted are the Exact Solution, G-ROM, EH-ROM, and SP-ROM. Note that only the SP-ROM produces reasonable energy and entropy results.

5.2. **A damped thermoelastic rod.** More useful from a ROM perspective is when the systems under consideration represent discretizations of infinite-dimensional metriplectic dynamics. Consider an infinite-dimensional example of the system in [36, Section 3.1] (specialized in [37]), where a 1-D elastic rod with coordinate $s \in [0, \ell]$ evolves as a damped Hamiltonian system with friction. The dynamics of this motion are governed by the metriplectic system

$$\begin{pmatrix} \dot{q} \\ \dot{p} \\ \dot{S} \end{pmatrix} = \begin{pmatrix} \frac{p}{m} \\ V'(q) - \gamma \frac{p}{m} \\ \gamma \frac{p^2}{m^2} \end{pmatrix} = \mathbf{L} \nabla E(q,p,e) + \mathbf{M}(q,p,e) \nabla S,$$

where V is a given potential function, γ is a constant controlling the rate of dissipation, $\nabla S, \nabla E$ now denote L^2-gradients, $m = m(s)$ is the mass density, $q = q(s)$ is the position, $p = p(s)$ is the momentum, and $e = e(s)$ is the internal energy representing the conversion of mechanical energy into heat. Explicitly, the state of the system can be described by $\mathbf{x}(s) = \begin{pmatrix} q(s) & p(s) & S(s) \end{pmatrix}^\intercal$ where the functions E and S are given by

$$E(p,q,e) = H(p,q) + S(e) = \int_0^\ell \left(\frac{p(s)^2}{2m(s)} + V\left(q(s)\right) \right) + \int_0^\ell e(s),$$

and where $H(p,q)$ is the Hamiltonian function for the system. Notice that H is no longer a conserved quantity but has been replaced by E which balances energetic and entropic contributions. A simple calculation using the definition $dF(v) = (\nabla F, v)_{L^2}$ yields the gradients and operators which describe these metriplectic dynamics: it follows that $\nabla S = \begin{pmatrix} 0 & 0 & 1 \end{pmatrix}^\top$,

$$\mathbf{L} = \begin{pmatrix} 0 & 1 & 0 \\ -1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \quad \mathbf{M} = \gamma \begin{pmatrix} 0 & 0 & 0 \\ 0 & 1 & -\frac{p}{m} \\ 0 & -\frac{p}{m} & \left(\frac{p}{m}\right)^2 \end{pmatrix}, \quad \nabla E = \begin{pmatrix} V'(q) \\ \frac{p}{m} \\ 1 \end{pmatrix}.$$

Modeling this evolution requires an appropriate discretization of the continuous system above. Consider a semi-discretization in the rod parameter with constant mass density m, so that s is represented by the vector $\mathbf{s} \in \mathbb{R}^N$ where N is the number of discretization points. Then, the discretized state (also denoted $\mathbf{x}$) becomes a $(2N+1)$-vector which evolves according to

$$\begin{pmatrix} \dot{\mathbf{q}} \\ \dot{\mathbf{p}} \\ \dot{S} \end{pmatrix} = \begin{pmatrix} \frac{\mathbf{p}}{m} \\ \mathbf{V}'(\mathbf{q}) - \gamma \frac{\mathbf{p}}{m} \\ \gamma \frac{|\mathbf{p}|^2}{m^2} \end{pmatrix} = \mathbf{L} \nabla E(\mathbf{x}) + \mathbf{M}(\mathbf{x}) \nabla S,$$

where $\nabla S = \begin{pmatrix} \mathbf{0} & \mathbf{0} & 1 \end{pmatrix}^\intercal$,

$$\mathbf{L} = \begin{pmatrix} \mathbf{0}_{N\times N} & \mathbf{I} & \mathbf{0} \\ -\mathbf{I} & \mathbf{0}_{N\times N} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & 0 \end{pmatrix}, \quad \mathbf{M}(\mathbf{x}) = \gamma \begin{pmatrix} \mathbf{0}_{N\times N} & \mathbf{0}_{N\times N} & \mathbf{0} \\ \mathbf{0}_{N\times N} & \mathbf{I} & -\frac{\mathbf{p}}{m} \\ \mathbf{0} & -\frac{\mathbf{p}^\intercal}{m} & \left(\frac{|\mathbf{p}|}{m}\right)^2 \end{pmatrix}, \quad \nabla E = \begin{pmatrix} \mathbf{V}'(\mathbf{q}) \\ \frac{\mathbf{p}}{m} \\ 1 \end{pmatrix}.$$

Notice that $S^{2N+1} = E^{2N+1} = 1 \neq 0$ and $\mathbf{M}$ decomposes as

$$\mathbf{M} = \sum_{\alpha=1}^N \mathbf{m}^\alpha \otimes \mathbf{m}^\alpha, \qquad \mathbf{m}^\alpha = \sqrt{\gamma} \begin{pmatrix} \mathbf{0} & \mathbf{e}_\alpha & -\frac{p_\alpha}{m} \end{pmatrix}^\intercal,$$

so that $\hat{\boldsymbol{\xi}}, \hat{\boldsymbol{\zeta}}$ can be computed as

$$\hat{\boldsymbol{\xi}} = \hat{\mathbf{L}} \wedge \mathbf{U}^{2N+1}, \qquad \hat{\boldsymbol{\zeta}}\left(\tilde{\mathbf{x}}\right) = \sum_{\alpha=1}^N \mathbf{A}^\alpha\left(\tilde{\mathbf{x}}\right) \otimes \mathbf{A}^\alpha\left(\tilde{\mathbf{x}}\right), \qquad \mathbf{A}^\alpha\left(\tilde{\mathbf{x}}\right) = \mathbf{U}^\intercal \mathbf{m}^\alpha\left(\tilde{\mathbf{x}}\right) \wedge \mathbf{U}^{2N+1}.$$

Moreover, although $\mathbf{M}$ depends on the full-order state $\tilde{\mathbf{x}} \in \mathbb{R}^{2N+1}$, each $\mathbf{m}^\alpha$ is linear in the solution and so the online cost can be made independent of this number (although it will still depend on N, the number of terms in the sum). More precisely, notice that

$$\mathbf{M} = \mathbf{C}\mathbf{C}^\intercal, \qquad \mathbf{C} = \sqrt{\gamma} \begin{pmatrix} \mathbf{0}_{N\times N} & \boldsymbol{I} & -\frac{\mathbf{p}}{m}, \end{pmatrix}^\intercal$$

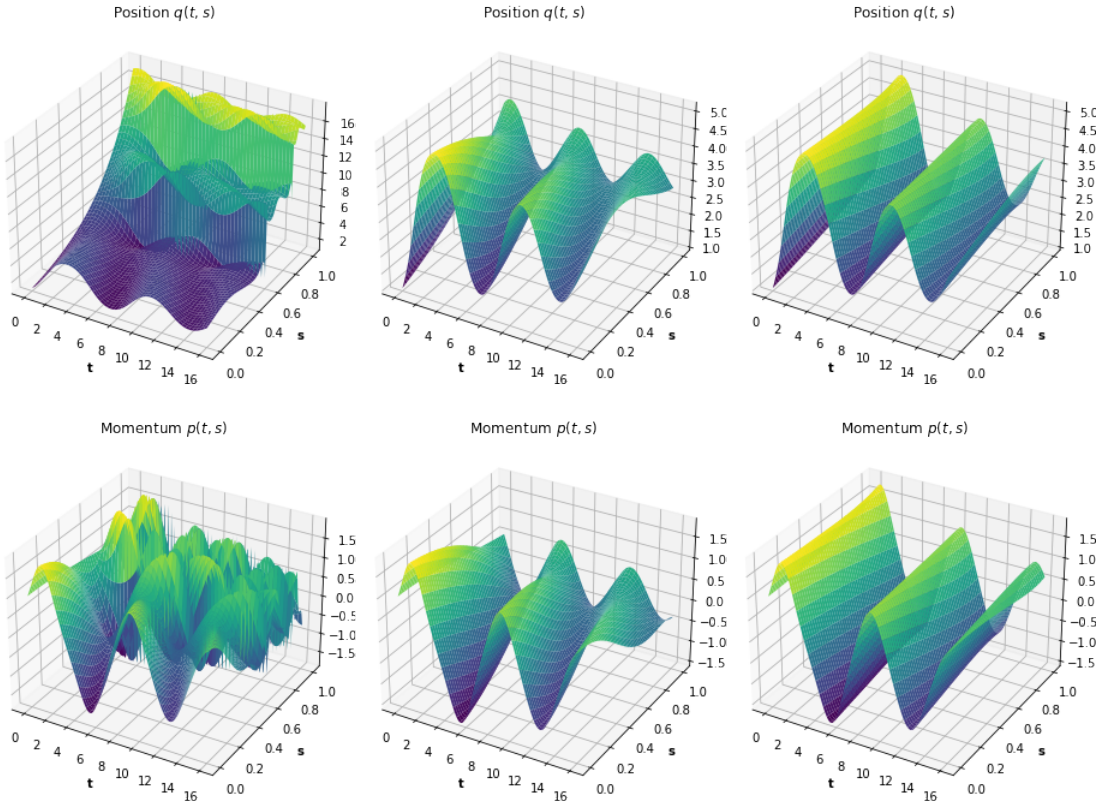

FIGURE 4. The position $q(t, s)$ and momentum $p(t, s)$ functions of three qualitatively different solutions to the thermoelastic rod problem contained in the training set.

so that $\mathbf{C}(\tilde{\mathbf{x}}) \in \mathbb{R}^{(2N+1)\times N}$ can be written as

$$\mathbf{C}(\tilde{\mathbf{x}}) = \mathbf{C}_0 + \mathbf{C}_1(\hat{\mathbf{x}}) = \sqrt{\gamma}\begin{pmatrix} \mathbf{0}_{N\times N} \\ \boldsymbol{I} \\ -\frac{\mathbf{p}^\intercal}{m} \end{pmatrix} + \frac{\sqrt{\gamma}}{m}\begin{pmatrix} \mathbf{0}_{N\times N} \\ \mathbf{0}_{N\times N} \\ -\left(\mathbf{U}^{N:2N}\hat{\mathbf{x}}\right)^\intercal \end{pmatrix},$$

where $\mathbf{U}^{N:2N}$ indicates the $N \times n$ matrix formed from rows N to $2N$ of $\mathbf{U}$. It follows that the reduced object

$$\mathbf{U}^\intercal\mathbf{C}(\tilde{\mathbf{x}}) = \mathbf{U}^\intercal\mathbf{C}_0 + \mathbf{U}^\intercal\mathbf{C}_1(\hat{\mathbf{x}}) = \mathbf{U}^\intercal\mathbf{C}_0 - \frac{\sqrt{\gamma}}{m}\mathbf{U}^{2N+1} \otimes \mathbf{U}^{N:2N}\hat{\mathbf{x}},$$

contains all $\mathbf{U}^\intercal\mathbf{m}^\alpha$ in its columns and requires only multiplication by $\hat{\mathbf{x}}$ online. This optimization has been used throughout this example on the FOM (where $\mathbf{p} = \mathbf{p}_0 + (\mathbf{p} - \mathbf{p}_0)$) as well as all ROMs. In addition, all simulations use the potential $V(q) = \cos q$ along with constants $\gamma = 8$, $\ell = 1$ and $N = 250$.

The performance of each ROM in this case is evaluated using a family of trajectories with parameterized initial conditions. More precisely, it is assumed that $\mathbf{x}_0$ satisfies the initial position and momentum conditions

$$\mathbf{q}_0(\mathbf{s}) = e^{\mu_1 \mathbf{s}}, \qquad \mu_1 \in [-0.2, 5.2],$$
$$\mathbf{p}_0(\mathbf{s}) = \frac{1}{1 + \mu_2 \mathbf{s}^2}, \qquad \mu_2 \in [-1, 1],$$

along with some initial entropy $S_0 \in [1, 3]$. Some representative solutions to the thermoelastic rod system with this type of initial data can be found in Figure 4. As in the previous example, 25 uniformly random instances of $\begin{pmatrix}\mu_1 & \mu_2 & S_0\end{pmatrix}^\intercal$ are drawn and used to form the initial conditions used for training the POD map $\mathbf{U}$. Note that the snapshot matrix $\mathbf{Y}$ includes snapshots of the shifted solution $\mathbf{x} - \mathbf{x}_0$ as well as the gradients $\nabla E(\mathbf{x})$ collected from the interval $[0, 8]$ in t-increments of 0.02. Again, the constant vector ∇S is

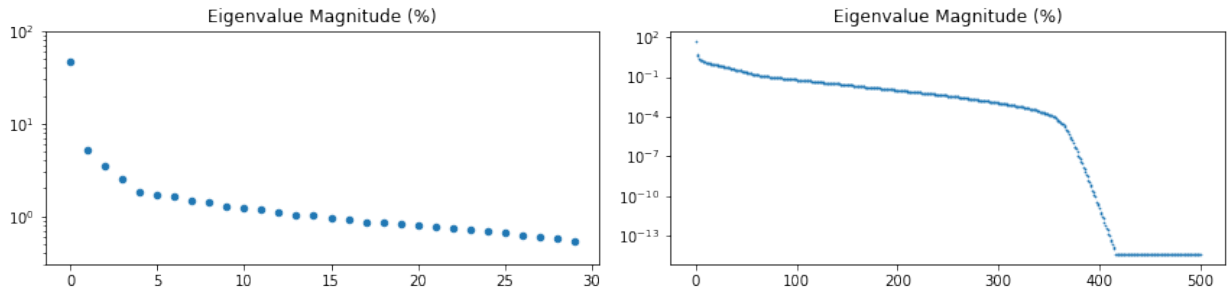

FIGURE 5. Eigenvalue plots corresponding to the snapshot matrix in the thermoelastic rod example. The y-axis displays the magnitude of each eigenvalue as a percentage of the total sum.

T	n	Method	$\mathcal{E}_r$ %	$\mathcal{E}_\infty$	$\lvert E(T)-E_0\rvert$	Time (s)	T	n	Method	$\mathcal{E}_r$ %	$\mathcal{E}_\infty$	$\lvert E(T)-E_0\rvert$	Time (s)
8	-	FOM	-	-	1.918×10^{-11}	0.1614	16	-	FOM	-	-	1.651×10^{-11}	0.2766
	10	SP-ROM	9.015	22.28	2.814×10^{-12}	0.1965		10	SP-ROM	8.166	26.49	1.648×10^{-12}	0.4123
		EH-ROM	4.896	7.980	6.687	0.06249			EH-ROM	4.134	7.980	36.43	0.1039
		G-ROM	8.336	21.65	16.99	0.1091			G-ROM	8.394	21.65	59.49	0.1975
	20	SP-ROM	4.020	7.938	1.108×10^{-12}	0.1762		20	SP-ROM	3.565	10.71	5.571×10^{-12}	0.4895
		EH-ROM	3.625	5.708	5.686	0.05898			EH-ROM	2.881	5.708	21.06	0.1107
		G-ROM	4.864	15.49	9.108	0.1186			G-ROM	5.304	15.49	61.26	0.1948
	40	SP-ROM	0.8734	0.5371	4.121×10^{-12}	0.2523		40	SP-ROM	0.9430	1.509	4.234×10^{-12}	0.4403
		EH-ROM	0.8767	0.5469	0.8563	0.07977			EH-ROM	0.9339	1.694	4.457	0.1378
		G-ROM	1.001	0.4779	1.555	0.1339			G-ROM	1.207	1.029	9.142	0.2258
48	-	FOM	-	-	1.253×10^{-11}	0.6041	96	-	FOM	-	-	3.439×10^{-12}	1.003
	10	SP-ROM	9.197	42.97	1.708×10^{-11}	0.6814		10	SP-ROM	10.11	44.97	8.413×10^{-12}	0.9792
		EH-ROM	18.20	154.0	280.6	0.2015			EH-ROM	92.49	846.9	1082	0.4351
		G-ROM	15.20	122.6	297.3	0.3931			G-ROM	72.28	787.7	1190	0.7518
	20	SP-ROM	4.456	21.09	2.643×10^{-12}	0.6928		20	SP-ROM	5.013	6.056	8.811×10^{-13}	0.8788
		EH-ROM	9.079	68.68	103.6	0.2628			EH-ROM	29.25	252.3	320.5	0.4126
		G-ROM	107.3	1498	3051	0.5186			G-ROM	-	-	-	-
	40	SP-ROM	1.302	5.538	4.832×10^{-13}	0.6735		40	SP-ROM	1.495	6.056	2.842×10^{-12}	1.427
		EH-ROM	2.063	14.42	21.07	0.2682			EH-ROM	5.369	41.49	49.38	0.5114
		G-ROM	3.908	32.51	67.35	0.5442			G-ROM	18.69	168.8	240.3	0.9306

TABLE 2. Results of the thermoelastic rod experiment. Dashes indicate "not applicable" when reporting the FOM, and lack of convergence when reporting the ROMs.

included as the last column of $\mathbf{Y}$. It is interesting to note that the first eigenvalue of the snapshot matrix is much larger than the rest, but the remaining eigenvalues decay relatively slowly until roughly $n = 350$ (see Figure 5). This indicates that these dynamics cannot be captured by only a few linear POD basis functions.

After training $\mathbf{U}$, the ROMs are evaluated online using the initial data corresponding to the parameter $\begin{pmatrix}0.65 & -0.1 & 1.9\end{pmatrix}^\intercal$ not included in the training set. The results of this procedure are tabulated in Table 5.2 and illustrated in Figures 6 and 7. Clearly, the greatest advantage of the metriplectic SP-ROM is its ability to preserve the energy conservation and entropy growth of the original system independently of n, leading to much greater accuracy and stability over time. Conversely, there appears to be little advantage to employing the more expensive SP-ROM over the cheaper, matrix-oriented EH-ROM when the integration takes place over small times, as there is not enough accumulation from the violation of the compatibility conditions (2) to significantly harm model performance. Note that the simple and straightforward G-ROM is unstable and inferior in almost every case. Since it has no knowledge of the internal structure of the system, it cannot accurately infer the original metriplectic dynamics.

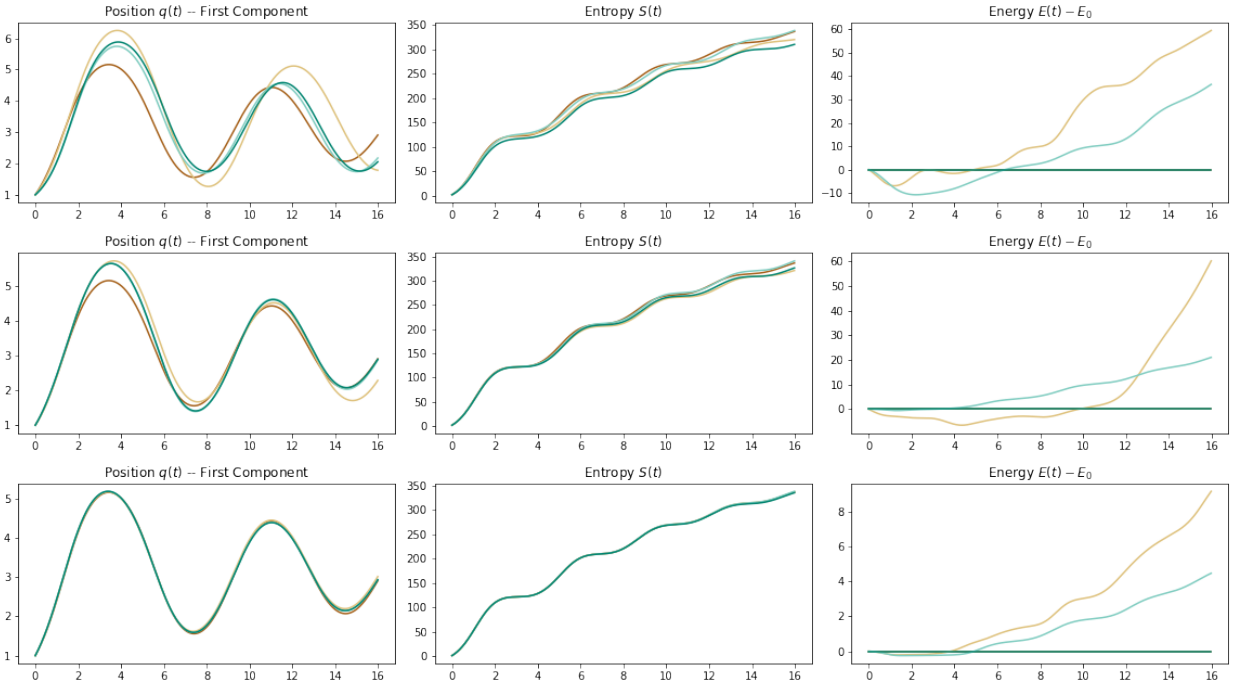

FIGURE 6. A comparison of ROM solutions for the 501-dimensional thermoelastic rod example when $T = 16$ and $n = 10, 20, 40$, respectively. Plotted are the Exact Solution, G-ROM, EH-ROM, and SP-ROM. Observe the convergence as predicted in Theorem 4.2.

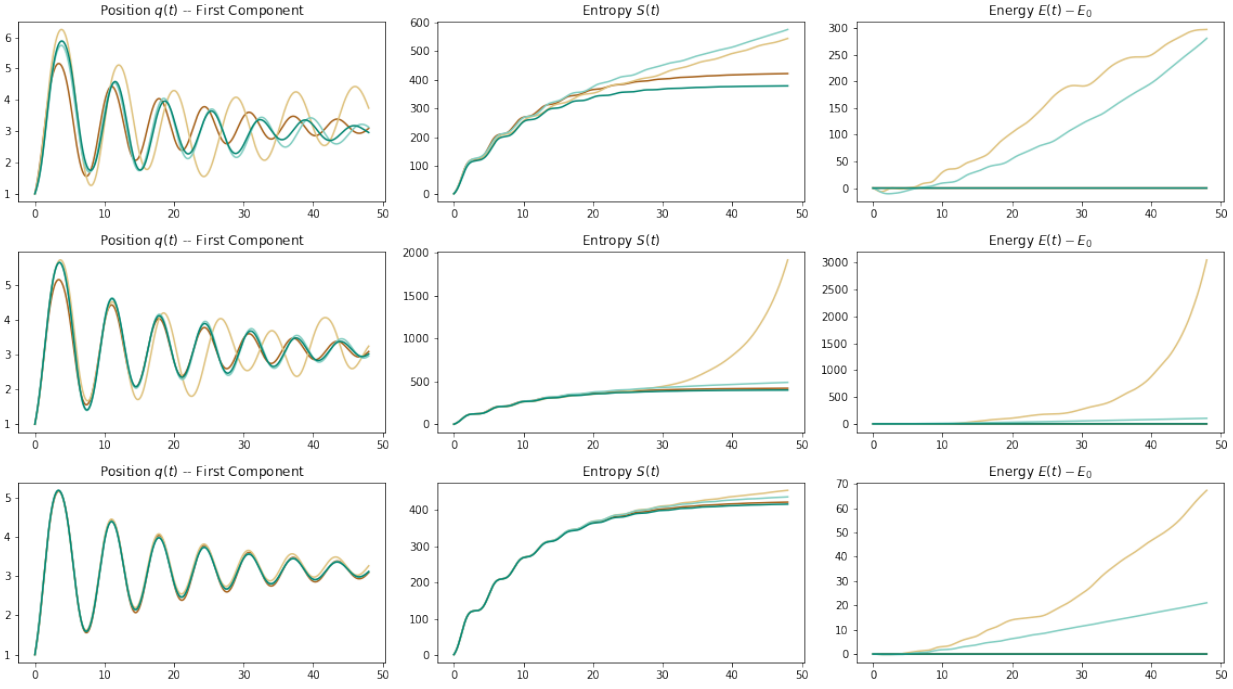

FIGURE 7. A comparison of ROM solutions for the 501-dimensional thermoelastic rod example when $T = 48$ and $n = 10, 20, 40$, respectively. Plotted are the Exact Solution, G-ROM, EH-ROM, and SP-ROM. Note that only the SP-ROM produces reasonable energy and entropy results.

$\mathcal{E}_r$ %			T						
			8	16	32	64	128	256	512
n	5	SP-ROM	19.28	17.96	17.18	17.44	17.55	17.36	17.18
		EH-ROM	11.36	10.02	12.34	668.0	-	-	-
		G-ROM	10.21	11.69	12.88	70.01	322.4	-	-
	10	SP-ROM	9.015	8.166	8.425	9.672	10.29	10.51	10.59
		EH-ROM	4.896	4.134	7.815	34.64	160.6	-	-
		G-ROM	8.336	8.394	8.409	27.23	509.0	-	-
	20	SP-ROM	4.020	3.565	3.952	4.750	5.118	5.231	5.262
		EH-ROM	3.624	2.881	4.858	14.54	51.36	922.8	-
		G-ROM	4.864	5.304	10.77	-	-	-	-

TABLE 3. Long-time integration results for the thermoelastic rod example. Values are the relative error in percentage form, and dashes indicate when the solver does not converge.

It is also useful to investigate the long-term stability of these ROMs. As the EH-ROM and the G-ROM are not truly metriplectic, it is expected that their performance will decay as the interval of integration moves far away from the training data. This is tested using the same experiment as above by varying the right endpoint of the temporal integration (recall that $\mathbf{U}$ is trained only on snapshots coming from the interval $[0, 8]$). Table 5.2 shows the results of integrating over ranges $[0, T]$ where $T = 2^k, 3 \leq k \leq 9$. As expected, the metriplectic SP-ROM is quite stable, while the others eventually break down. It is interesting that the naive G-ROM is unpredictable, exhibiting better stability when $n = 10$ than when $n = 20$. Note that the SP-ROM is useful for reducing computational costs in this regime also, as the time necessary to integrate the FOM at $T = 512$ is roughly 5.5 seconds versus 2-3 seconds (depending on n) for the SP-ROM.

6. Conclusion

A new strategy for the model reduction of metriplectic systems has been proposed and shown to guarantee a strong form of the first and second laws of thermodynamics. By preserving the original metriplectic structure at the algebraic level, the proposed ROM is able to produce more realistic energy and entropy profiles than other POD-ROMs which cannot guarantee structure-preservation. It has been shown that the metriplectic POD-ROM conserves energy to arbitrary precision regardless of the reduced dimension and converges to the true solution as this dimension increases. As the proposed ROM is trained similarly to standard POD-ROMs, it is useful as a drop-in replacement for metriplectic problems demanding physically realistic solutions where conserved quantities are important or longer time scales where stability is relevant. Future work will investigate applications to more complex problems such as those mentioned in Section 1.2, as well as effective methods such as Discrete Empirical Interpolation for making the metriplectic ROM completely independent of the full-order dimension. As several important problems in fluid mechanics are also known to have a metriplectic form (see e.g. [1]), it is especially interesting to consider the application of these techniques to realistic oceanic and atmospheric models which require strict adherence to physical laws.

Acknowledgments

This work is partially supported by U.S. Department of Energy Scientific Discovery through Advanced Computing under grants DE-SC0020270 and DE-SC0020418.

References

[1] H. C. Öttinger, "Nonequilibrium thermodynamics for open systems," *Phys. Rev. E*, vol. 73, p. 036126, Mar 2006.
[2] P. J. Morrison, "Some observations regarding brackets and dissipation," *Center for Pure and Applied Mathematics Report PAM-228, University of California, Berkeley*, 1984.
[3] M. Grmela and H. C. Öttinger, "Dynamics and thermodynamics of complex fluids. i. development of a general formalism," *Physical Review E*, vol. 56, no. 6, p. 6620, 1997.
[4] Y. Suzuki, "A GENERIC formalism for Korteweg-type fluids: I. a comparison with classical theory," *Fluid Dynamics Research*, vol. 52, no. 1, p. 015516, 2020.

[5] N. J. Wagner, "The Smoluchowski equation for colloidal suspensions developed and analyzed through the GENERIC formalism," *Journal of non-Newtonian fluid mechanics*, vol. 96, no. 1-2, pp. 177–201, 2001.
[6] A. Ait-Kadi, A. Ramazani, M. Grmela, and C. Zhou, ""volume preserving" rheological models for polymer melts and solutions using the GENERIC formalism," *Journal of Rheology*, vol. 43, no. 1, pp. 51–72, 1999.
[7] M. Materassi and E. Tassi, "Metriplectic framework for dissipative magneto-hydrodynamics," *Physica D: Nonlinear Phenomena*, vol. 241, no. 6, pp. 729–734, 2012.
[8] M. Materassi and P. J. Morrison, "Metriplectic torque for rotation control of a rigid body," *Cybernetics and Physics*, 2018.
[9] C. Caligan and C. Chandre, "Conservative dissipation: How important is the Jacobi identity in the dynamics?," *Chaos: An Interdisciplinary Journal of Nonlinear Science*, vol. 26, no. 5, p. 053101, 2016.
[10] M. H. Duong, M. A. Peletier, and J. Zimmer, "GENERIC formalism of a Vlasov–Fokker–Planck equation and connection to large-deviation principles," *Nonlinearity*, vol. 26, no. 11, p. 2951, 2013.
[11] P. Betsch and M. Schiebl, "Energy-momentum-entropy consistent numerical methods for large-strain thermoelasticity relying on the GENERIC formalism," *International Journal for Numerical Methods in Engineering*, vol. 119, no. 12, pp. 1216–1244, 2019.
[12] I. Romero, "Algorithms for coupled problems that preserve symmetries and the laws of thermodynamics: Part i: Monolithic integrators and their application to finite strain thermoelasticity," *Computer Methods in Applied Mechanics and Engineering*, vol. 199, no. 25-28, pp. 1841–1858, 2010.
[13] K. Lee, N. Trask, and P. Stinis, "Machine learning structure preserving brackets for forecasting irreversible processes," *Advances in Neural Information Processing Systems*, vol. 34, 2021.
[14] Z. Zhang, Y. Shin, and G. E. Karniadakis, "GFINNs: GENERIC formalism informed neural networks for deterministic and stochastic dynamical systems," *arXiv preprint arXiv:2109.00092*, 2021.
[15] L. Peng and K. Mohseni, "Symplectic model reduction of Hamiltonian systems," *SIAM Journal on Scientific Computing*, vol. 38, no. 1, pp. A1–A27, 2016.
[16] Y. Gong, Q. Wang, and Z. Wang, "Structure-preserving Galerkin POD reduced-order modeling of Hamiltonian systems," *Computer Methods in Applied Mechanics and Engineering*, vol. 315, pp. 780–798, 2017.
[17] B. M. Afkham and J. S. Hesthaven, "Structure preserving model reduction of parametric Hamiltonian systems," *SIAM Journal on Scientific Computing*, vol. 39, no. 6, pp. A2616–A2644, 2017.
[18] B. Maboudi Afkham and J. S. Hesthaven, "Structure-preserving model-reduction of dissipative Hamiltonian systems," *Journal of Scientific Computing*, vol. 81, no. 1, pp. 3–21, 2019.
[19] K. C. Sockwell, *Mass Conserving Hamiltonian-Structure-Preserving Reduced Order Modeling for the Rotating Shallow Water Equations Discretized by a Mimetic Spatial Scheme*. PhD thesis, The Florida State University, 2019.
[20] R. V. Polyuga and A. Van der Schaft, "Structure preserving model reduction of port-Hamiltonian systems by moment matching at infinity," *Automatica*, vol. 46, no. 4, pp. 665–672, 2010.
[21] C. Beattie and S. Gugercin, "Structure-preserving model reduction for nonlinear port-Hamiltonian systems," in *2011 50th IEEE conference on decision and control and European control conference*, pp. 6564–6569, IEEE, 2011.
[22] S. Gugercin, R. V. Polyuga, C. Beattie, and A. Van Der Schaft, "Structure-preserving tangential interpolation for model reduction of port-Hamiltonian systems," *Automatica*, vol. 48, no. 9, pp. 1963–1974, 2012.
[23] S. Chaturantabut, C. Beattie, and S. Gugercin, "Structure-preserving model reduction for nonlinear port-Hamiltonian systems," *SIAM Journal on Scientific Computing*, vol. 38, no. 5, pp. B837–B865, 2016.
[24] B. Liljegren-Sailer, "On port-Hamiltonian modeling and structure-preserving model reduction," 2020.
[25] Z. Bai and R.-C. Li, "Structure-preserving model reduction using a krylov subspace projection formulation," *Communications in Mathematical Sciences*, vol. 3, no. 2, pp. 179–199, 2005.
[26] S. Lall, P. Krysl, and J. E. Marsden, "Structure-preserving model reduction for mechanical systems," *Physica D: Nonlinear Phenomena*, vol. 184, no. 1-4, pp. 304–318, 2003.
[27] C. Beattie and S. Gugercin, "Interpolatory projection methods for structure-preserving model reduction," *Systems & Control Letters*, vol. 58, no. 3, pp. 225–232, 2009.
[28] H. Egger, T. Kugler, B. Liljegren-Sailer, N. Marheineke, and V. Mehrmann, "On structure-preserving model reduction for damped wave propagation in transport networks," *SIAM Journal on Scientific Computing*, vol. 40, no. 1, pp. A331–A365, 2018.
[29] Y. LIANG, H. LEE, S. LIM, W. LIN, K. LEE, and C. WU, "Proper orthogonal decomposition and its applications—part i: Theory," *Journal of Sound and Vibration*, vol. 252, no. 3, pp. 527–544, 2002.
[30] D. Hestenes and G. Sobczyk, *Clifford algebra to geometric calculus: a unified language for mathematics and physics*, vol. 5. Springer Science & Business Media, 2012.
[31] L. W. Tu, *Differential geometry: connections, curvature, and characteristic classes*, vol. 275. Springer, 2017.
[32] S. Chaturantabut and D. C. Sorensen, "Nonlinear model reduction via discrete empirical interpolation," *SIAM Journal on Scientific Computing*, vol. 32, no. 5, pp. 2737–2764, 2010.
[33] P. Virtanen, R. Gommers, T. E. Oliphant, M. Haberland, T. Reddy, D. Cournapeau, E. Burovski, P. Peterson, W. Weckesser, J. Bright, S. J. van der Walt, M. Brett, J. Wilson, K. J. Millman, N. Mayorov, A. R. J. Nelson, E. Jones, R. Kern, E. Larson, C. J. Carey, İ. Polat, Y. Feng, E. W. Moore, J. VanderPlas, D. Laxalde, J. Perktold, R. Cimrman, I. Henriksen, E. A. Quintero, C. R. Harris, A. M. Archibald, A. H. Ribeiro, F. Pedregosa, P. van Mulbregt, and SciPy 1.0 Contributors, "SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python," *Nature Methods*, vol. 17, pp. 261–272, 2020.
[34] A. C. Hindmarsh, "ODEPACK, a systematized collection of ODE solvers," *Scientific computing*, pp. 55–64, 1983.

[35] X. Shang and H. C. Öttinger, "Structure-preserving integrators for dissipative systems based on reversible-irreversible splitting," *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences*, vol. 476, no. 2234, p. 20190446, 2020.
[36] A. Mielke, "Formulation of thermoelastic dissipative material behavior using GENERIC," *Continuum Mechanics and Thermodynamics*, vol. 23, no. 3, pp. 233–256, 2011.
[37] R. C. Kraaij, A. Lazarescu, C. Maes, and M. Peletier, "Fluctuation symmetry leads to GENERIC equations with non-quadratic dissipation," *Stochastic Processes and their Applications*, vol. 130, no. 1, pp. 139–170, 2020.

Email address: agruber@fsu.edu, mgunzburger@fsu.edu, ju@math.sc.edu, wangzhu@math.sc.edu

[1] Department of Scientific Computing, Florida State University, 400 Dirac Science Library, Tallahassee, FL 32306, USA

[2] Department of Mathematics, University of South Carolina, 1523 Greene Street, Columbia, SC 29208, USA